NUCLEOTIDE SEQUENCES FOR GENE REGULATION AND METHODS OF USE THEREOF

This work was made, in part, with Government support by the National Institutes of Health grant numbers HL38644 and HL53713. The Government has certain rights in this invention.

FIELD OF THE INVENTION

The invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In particular, the invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner and/or in a liver-specific manner. The invention further relates to methods of using the regulatory nucleic acid sequences provided herein for age-related and/or liver-specific expression of nucleotides sequences of interest. The invention also relates to host cells and to transgenic non-human animals which harbor the regulatory nucleic acid sequences of the invention. The compositions and methods of the invention are useful in regulating expression of a nucleotide sequence of interest in an age-related and/or liver-specific manner.

BACKGROUND OF THE INVENTION

A multitude of human diseases (*e.g.*, thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) are associated with increasing age and result in serious effects on the quality of life and on the life expectancy of individuals suffering from such diseases. Other diseases (*e.g.*, cirrhosis, primary and metastatic neoplasia, Wilson disease, hepachromatosis, infectious hepatitis, hepatic necrosis, Gilbert disease, Criggler-Najar disease) which afflict the liver also have serious clinical manifestations and are responsible for high morbidity and mortality.

The treatment of age-related diseases (*i.e.*, diseases whose prevalence and/or severity of clinical manifestations increases with the age of the patient) and diseases afflicting the liver focuses on the alleviation of the general symptoms of the disease using one or a combination of two modalities, *i.e.*, non-pharmacological treatment and pharmacological treatment. Non-pharmacological treatment include, for example, periods of bed rest and dietary changes. Non-pharmacological treatment is often used as an adjunct to

pharmacological treatment which involves the use of drugs. Unfortunately, many of the commonly used pharmacological agents have numerous side effects and their use is further exacerbated by the non-responsiveness by many patients with severe disease, who, paradoxically, are in most need of treatment. Both non-pharmacological and pharmacological treatments provide unsatisfactory approaches to treating age-related and liver-associated diseases because these approaches are often ineffective, their effects are inconsistent, and are directed to alleviating the general symptoms of disease, rather than to specifically addressing the source of morbidity and mortality. Moreover, no suitable animal models are currently available to rationally design drugs which target specific biochemical and physiological pathways which are associated with age-related and with liver-associated diseases.

What is needed are methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

SUMMARY OF THE INVENTION

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner, as well as nucleic acid sequences which direct liver-specific expression of a gene of interest. Further provided by the invention are transgenic animals which may be used as models for age-related and/or liver specific diseases.

In one embodiment, the invention provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to any particular type or source of nucleic acids sequence of interest, in a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be restricted to any particular type or source of promoter sequence, in an alternative preferred embodiment, the promoter sequence is selected from human factor IX

promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. It is not contemplated that the invention be limited to any particular age regulatory sequence which is a portion of SEQ ID NO:1. However, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:1 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention to any particular rage regulatory sequence which s a portion of SEQ ID NO:3, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

Also provided by the invention is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In another preferred embodiment, the host cell is a gamete. In yet another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without limiting the invention to the type or source of the nucleic acid sequence of interest, in one preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be limited to the type or source of the promoter sequence, in an alternative

preferred embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. Though it is not contemplated that the invention be limited to the portion of SEQ ID NO:1 which has age-related regulatory activity, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:1 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention the portion of SEQ ID NO:3 which has age-related regulatory activity, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58; SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

Also provided herein is a host cell containing recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative preferred embodiment, the host cell is a gamete. In another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a method, comprising: a) providing: i) a cell, ii) a nucleic acid sequence of interest, iii) a promoter sequence, and iv) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the one or more age regulatory sequences to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. Without intending to limit the treated cell to any particular environment, in one preferred embodiment, the treated cell is comprised in a tissue or organ in a living animal.

The invention further provides a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

Also provided by the present invention is a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the portion is SEQ ID NO:2. In an alternative embodiment, the portion is selected from SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37 and SEQ ID NO:38.

The invention also provides a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the portion is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:1 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 63% and less than 100% homology to the SEQ ID NO:1.

The invention also provides a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:3 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 60% and less than 100% homology to the SEQ ID NO:3.

The invention additionally provides a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

The invention also provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

The invention additionally provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) SEQ ID NO:1; and v) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the SEQ ID NO:1 and the SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell. In one embodiment, the cell expresses a recombinant protein identified as SEQ ID NO:47. In an alternative embodiment, the cell is selected from HepG2 cell, fibroblast cell, myoblast cell, and endothelial cell. In another embodiment, the cell is a fertilized egg cell, and the transgenic cell is a transgenic fertilized egg cell. In a preferred embodiment, the method further comprises d) introducing the transgenic fertilized egg cell into a non-human animal and permitting the animal to deliver progeny containing the transgene. In a more preferred embodiment, the progeny is characterized by age-related expression of the nucleic acid sequence of interest. In an alternative more preferred embodiment, the progeny is characterized by liver-specific expression of the nucleic acid sequence of interest. In another preferred embodiment, the fertilized egg cell is derived from a mammal of the order Rodentia. In a more preferred embodiment, the fertilized egg cell is a mouse fertilized egg cell. In yet another embodiment, the promoter is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK

promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase.

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) a portion of SEQ ID NO:1; and v) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the portion of SEQ ID NO:1 and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Additionally provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Also provided herein is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

The invention further provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Further provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

The invention further also provides additional sequences derived from the hFIX gene. In particular, the invention provides a substantially purified nucleic acid sequence comprising at least a portion of SEQ ID NO:93. In one embodiment, the portion has age-related regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a preferred embodiment, the portion is SEQ ID NO:91.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under high stringency hybridization conditions with a nucleotide sequence selected from SEQ ID NO:91, the complement of SEQ ID NO:91, SEQ ID NO:93, and the complement of SEQ ID NO:93. In one embodiment, the nucleic acid sequence has age-related regulatory activity.

The invention also provides a substantially purified nucleic acid sequence comprising a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof.

The invention further provides expression vectors containing hFIX sequences. In one embodiment, the invention provides a recombinant expression vector comprising in operable

combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodimnt, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another alternative embodiment, the portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a more preferred embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

The invention further provides cells containing hFIX-derived sequences. In particular, the invention provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In another alternative embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention additionally provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In yet a further embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

Also provided herein is host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In yet

another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hFIX-derived sequences. For example, the invention provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the age-related regulatory sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

The invention also provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

In addition to sequences from the hFIX gene, the invention provides sequences derived from the hPC gene. In particular, the invention provides substantially purified nucleic acid sequence comprising a nucleotide sequence selected from at least a portion of SEQ ID NO:85, and at least a portion of SEQ ID NO:92. In one embodiment, the portion has activity selected from age-related regulatory activity and regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID

NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In another alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with a nucleotide sequence selected from SEQ ID NO:92, the complement of SEQ ID NO:92, SEQ ID NO:85, the complement of SEQ ID NO:85, SEQ ID NO:89, the complement of SEQ ID NO:89, SEQ ID NO:90, and the complement of SEQ ID NO:90. In one embodiment, the nucleic acid sequence has activity selected from age-related regulatory activity and regulatory activity.

The invention additionally provides a substantially purified nucleic acid sequence comprising a functional homolog of a sequence having activity selected from age-related regulatory activity and regulatory activity, the sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90.

Also provided by the invention is a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the expression vector of Claim 38, wherein the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion is

selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In a further embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

The invention also provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In a further embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin,

protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter, T_3 promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In an alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

Further provided herein is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In yet another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hPC sequences. In particular, the invention discloses a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the nucleotide sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

Also provided herein is a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the structure of eleven exemplary human FIX minigene expression constructs and relative *in vitro* transient expression activities (ng hFIX/10^6 cells/48 hr).

Figure 2 shows graphs of longitudinal analyses of transgenic mice which carry -416FIXm1 (A), -416FIXm1/1.4 (B), -590FIXm1 (C), -679FIXm1 (D), and -770FIXm1 (E) expression vectors and which produce high initial prepubertal, but rapidly decreasing, hFIX expression levels with age.

Figure 3 shows a Northern blot of human FIX mRNA levels (A) and a gel showing hFIX transgene DNA levels as determined by multiplex PCR analysis (B) in the livers and tails of animals carrying -416FIXm1.

Figure 4 shows graphs of longitudinal analysis of transgenic mice which carry -802FIXm1 (A), -802FIXm1/1.4 (B), -2231FIXm1 (C), -2231FIXm1/1.4 (D) and -416FIXm1/AE5' (E) expression vectors and which produce hFIX at stable and increasing levels with age.

Figure 5 shows a Northern blot of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 expression vectors.

Figure 6 is a gel of a gel electrophoretic mobility shift assay using mouse liver nuclear extract (NEs) from three different age groups, and using double-stranded oligonucleotides containing a PEA-3 element nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (A), and using a competition assay for ^{32}P-labelled double stranded oligonucleotides containing the PEA-3 nucleotide sequence (B).

Figure 7 is a Northern blot showing tissue specificity of hFIX expression in transgenic mice carrying -416FIXm1 (A) and -802 FIXm1 (B) expression vectors.

Figure 8A-E shows the nucleotide sequence (SEQ ID NO:4) of, and eight amino acid sequences (SEQ ID NOs:5 to 12) which together form, the human factor IX (GenBank accession number K02402). The initiation transcription site (nucleotide 1) and the poly-A addition site (nucleotide 32,757) are identified by solid circles. The solid vertical arrows indicate the intron-exon splice junction. The five Alu repetitive sequences have been underlined, while the 5-base insert in intron A and the AATAAA sequence in exon VIII are boxed. The cleavage or termination site at the 3' end of the gene (CATTG) is underlined with a dashed line.

Figure 9 shows the cDNA sequence (SEQ ID NO:13) (A) and encoded polypeptide sequence (SEQ ID NO:47) (B) of mouse PEA-3 (GenBank accession number X63190).

Figure 10 A-D shows the cDNA sequence (SEQ ID NO:42) of the human α1-antitrypsin gene (GenBank accession number K02212).

Figure 11 shows the DNA sequence (SEQ ID NO:43) of human antithrombin III (GenBank accession number A06100).

Figure 12 shows the cDNA sequence (A) (SEQ ID NO:49) (GenBank accession number X02750) and genomic DNA sequence (B) (SEQ ID NO:50) (GenBank accession number M11228) of human protein C.

Figure 13 (A-E) shows the nucleic acid sequences (SEQ ID NOs:76-83) of exemplary homologs of AE3' (SEQ ID NO:3).

Figure 14 shows the nucleotide sequence (from nt -1462 to nt +1; SEQ ID NO:85) which is located at the 5'-end of the human protein C gene.

Figure 15 shows the structure of eight exemplary human protein C minigene expression constructs.

Figure 16 shows the relative *in vitro* transient expression activities for five exemplary human protein C minigene expression constructs.

Figure 17 shows graphs of longitudinal analyses of transgenic mice which carry -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3' (C) expression vectors.

DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural state, or when obtained from its actual source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising SEQ ID NO:1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain (at a minimum) at least a portion of the sense or coding strand (*i.e.*, the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (*i.e.*, the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An

"isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated.

The term "recombinant" when made in reference to a DNA sequence refers to a DNA sequence which is comprised of segments of DNA joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a polypeptide sequence refers to a polypeptide sequence which is expressed using a recombinant DNA sequence.

As used herein, the terms "vector" and "vehicle" are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a tissue or animal refers to a tissue or animal, respectively, which comprises one or more cells that contain a transgene, or whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and animals may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

A "non-human animal" refers to any animal which is not a human and includes vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia. The term "order Rodentia" refers to rodents *i.e.*, placental mammals (class Euthria) which include the family Muridae (*e.g.*, rats and mice), most preferably mice.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (*e.g.*, treat disease, confer improved qualities, *etc.*), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (*e.g.*, reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, *etc.*), and non-coding regulatory sequences which do not encode an mRNA or protein product (*e.g.*, promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, *etc.*).

As used herein, the terms "complementarity," or "complementary" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (*i.e.*, partial identity) or complete homology (*i.e.*, complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (*i.e.*, an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific

binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." [Coombs J (1994) *Dictionary of Biotechnology*, Stockton Press, New York NY]. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\%\ G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl [*see e.g.*, Anderson and Young, Quantitative Filter Hybridization, *in Nucleic Acid Hybridization* (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

Low stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 1% SDS, 5X Denhardt's reagent [50X Denhardt's contains the following per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml

denatured salmon sperm DNA followed by washing in a solution comprising 0.2X SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE, 1% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

Those skilled in the art know that whereas higher stringencies may be preferred to reduce or eliminate non-specific binding of the nucleotide sequence of SEQ ID NOs:1 or 3 with other nucleic acid sequences, lower stringencies may be preferred to detect a larger number of nucleic acid sequences having different homologies to the nucleotide sequence of SEQ ID NOs:1 and 3.

As used herein, the terms "regulatory element" and "regulatory sequence" interchangeably refer to a nucleotide sequence which does not encode RNA or a protein and which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, *etc.* In contrast, the term "regulatory gene" refers to a DNA sequence which encodes RNA or a protein (*e.g.*, transcription factor) that controls the expression of other genes.

Regulatory elements may be tissue specific or cell specific. The term "tissue specific" as it applies to a regulatory element refers to a regulatory element that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (*e.g.*, liver) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (*e.g.*, lung).

Tissue specificity of a regulatory element may be evaluated by, for example, operably linking a reporter gene to a promoter sequence (which is not tissue-specific) and to the regulatory element to generate a reporter construct, introducing the reporter construct into the genome of an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (*e.g.*, detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic animal. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the regulatory element is "specific" for the tissues in which greater levels of expression are detected. Thus, the term "tissue-specific" (*e.g.*, liver-specific) as used herein is a relative term that does not require absolute specificity of expression. In other words, the term "tissue-specific" does not require that one tissue have extremely high levels of expression and another tissue have no expression. It is sufficient that expression is greater in one tissue than another. By contrast, "strict" or "absolute" tissue-specific expression is meant to indicate expression in a single tissue type (*e.g.*, liver) with no detectable expression in other tissues.

The term "cell type specific" as applied to a regulatory element refers to a regulatory element which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide

sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a regulatory element also means a regulatory element capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue.

Cell type specificity of a regulatory element may be assessed using methods well known in the art, *e.g.*, immunohistochemical staining and/or Northern blot analysis. Briefly, for immunohistochemical staining, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is regulated by the regulatory element. A labeled (*e.g.*, peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (*e.g.*, with avidin/biotin) by microscopy. Briefly, for Northern blot analysis, RNA is isolated from cells and electrophoresed on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (*i.e.*, upstream) of a nucleotide sequence of interest whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (*e.g.*, heat shock, chemicals, *etc.*). In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (*e.g.*, heat shock, chemicals, *etc.*) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The terms "essentially consisting of" and "consisting essentially of" are equivalent terms, and when in reference to a nucleic acid sequence they are intended to refer to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in the nucleic acid sequence, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in the nucleic acid sequence, and in which the biological activity of the nucleotide sequences is from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the biological activity of the nucleic acid sequence. To illustrate, the term "a nucleic acid sequence consisting essentially of SEQ ID NO:1" refers to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in SEQ ID NO:1, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in SEQ ID NO:1, and in which the nucleotide sequences exhibit from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the age-related regulatory activity, and/or of the liver-specific activity of SEQ ID NO:1.

A "functional homolog" of a nucleotide sequence which is derived from the hFIX gene shown in Figure 8 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hFIX-derived nucleotide sequence (*i.e.*, with the entire, or a portion of the, hFIX sequence of Figure 8), and which has age-related regulatory activity and/or liver-specific activity. For example, a functional homolog of SEQ ID NO:33 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:33, and which have age-related regulatory activity and/or liver-specific activity.

A "functional homolog" of a nucleotide sequence which is derived from the hPC sequence shown in Figure 14 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hPC-derived nucleotide sequence (*i.e.*, with the entire, or a portion of the, hPC sequence of Figure 14), and which has age-related regulatory activity and/or regulatory activity. For example, a functional homolog of SEQ ID NO:93 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:93, and which have age-related regulatory activity and/or regulatory activity.

DESCRIPTION OF THE INVENTION

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In on embodiment, the invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner. Yet more particularly, the exemplary age-regulatory element 5' (AE5') has been discovered to regulate stable gene expression over time *in vivo*, while the exemplary age-regulatory element 3' (AE3') has been discovered to regulate increased gene expression over time *in vivo*. In another embodiment, the invention provides nucleic acid sequences (*e.g.*, AE5') which direct liver-specific expression of a gene of interest. In yet another embodiment, the invention provides transgenic animals which harbor the nucleic acid sequences provided herein and which expres a nucleotide sequence of interest in an age-related and/or liver-specific manner. The nucleic acid sequences provided herein are useful in, for example, identifying and isolating functional homologs of AE5' and AE3', and amplifying at least a portion of AE5' and AE3'. Importantly, the nucleic acid sequences of the invention are also useful in age-related expression and/or liver-specific expression of a nucleotide sequence of interest in an animal, in gene therapy, and in reducing expression of factor IX in an animal.

In addition to regulatory sequences which are derived from the human factor IX gene (*e.g.*, AE5', AE3', and AE3''), the invention further provides sequences which are derived from the human protein C (hPC) gene and which are characterized by age-related regulatory activity and regulatory activity.

The invention is further discussed under (A) Regulatory Nucleic Acid Sequences, (B) Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (C) Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (D) Methods For Regulating Gene Expression, (E) Gene Therapy, and (F) Reducing Expression Of Factor IX In An Animal.

A. Regulatory Nucleic Acid Sequences

The invention provides regulatory sequences which are derived from the hFIX and hPC genes.

i. Regulatory Nucleic Acid Sequences From The hFIX Gene

The regulatory nucleic acid sequences of the invention and their surprising properties in regulating gene expression were discovered during the inventor's investigation of the mechanisms underlying age-associated regulation of the human factor IX, which is involved in blood coagulation. Blood coagulation plays a critical role not only in homeostasis, but also in many physiological and pathological conditions [Saito in *Disorders of Hemostasis*, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi *et al.* (1993) Blood Coagul. Fibrinol. 4:953-974]. Blood coagulation potential in humans as well as in other mammals reaches the young adult level around the age of weaning [Yao *et al.* (1991) Thromb. Haemost. 65:52-58; Andrew *et al.* (1992) Blood 80:1998-2005; Andrew *et al.* Blood (1987) 70:165-172; Andrew *et al.* (1988) Blood 72:1651-1657]. This is followed by a gradual increase in coagulation potential during young adulthood, and an almost two-fold increase by old age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi *et al.* (1996) Thromb. Haemost. 76:965-969]. This age-associated increase in coagulation potential takes place in healthy centenarians [Marie *et al.* (1995) Blood 85:3144-3149], indicating that the increase is a normal phenomenon associated with aging.

It is the inventors' consideration that this increase in coagulation potential may make a crucial contribution to the development and progression of age-associated diseases such as cardiovascular and thrombotic disorders [Conlan *et al.* (1993) The Atherosclerosis Risk in Communities (ARIC) Study 70:380-385; Balleisen *et al.* (1985) Thromb. Haemost. 54:475-479; Rode *et al.* (1996) Nat. Med. 2:293-298; Woodward *et al.* (1997) Brit. J. Haemat. 97:785-797]. The inventors' consideration was based on the observation that this increase in blood coagulation potential coincides with plasma level increases of pro-coagulant factors such as factor IX, whereas plasma levels of anti-coagulation factors (such as antithrombin III and protein C) or of factors involved in fibrinolysis are only marginally affected [Conlan *et al.* (1994) The Atherosclerosis Risk in Committees (ARIC) Study 72:551-556; Lowe *et al.* (1997) Brit. J. Haemat. 97:775-784]. These facts strongly suggested to the inventors that the

observed increase in blood coagulation activity with advancing age is due to regulated events. Plasma levels of each protein factor involved in blood coagulation, fibrinolysis and their regulatory systems are presumably determined by the balance of the many processes involved. At present, little is known about why an advancing age-associated increase in blood coagulation activity exists, or what molecular mechanisms are involved in age-dependent regulation (homeostasis) of blood coagulation [Finch in *Longevity, Senescence, and the Genome*, The University of Chicago Press, Chicago, 1990].

Blood coagulation factor IX (FIX) occupies a key position in the blood coagulation cascade where the intrinsic and extrinsic pathways merge [Saito in *Disorders of Hemostasis*, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi *et al.* (1993) Blood Coagul. Fibrinol. 4:953-974]. FIX is synthesized in the liver with strict tissue-specificity, and its deficiency results in the bleeding disorder hemophilia B. In normal humans the plasma activity and protein concentration levels of human FIX (hFIX) increase with advancing age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi *et al.* (1996) Thromb. Haemost. 76:965-969]. Mouse FIX (mFIX) plasma activity also increases with age in a manner similar to hFIX, and is directly correlated with an increase in liver mFIX messenger RNA (mRNA) level [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi *et al.* (1996) Thromb. Haemost. 76:965-969]. However, nothing else is known about the molecular mechanisms underlying such an increase. In investigating the basic molecular mechanisms responsible for age-associated regulation of hFIX, the inventors discovered the nucleotide sequences which regulate age-associated expression, and which direct liver-specific expression, of the exemplary hFIX gene.

The discovery of the invention sequences was made possible, in part, by the inventors' use of the hFIX promoter in combination with the coding sequence for hFIX instead of with the coding sequence for commonly used reporter proteins. The discovery of the surprising functions of the nucleotide sequences provided herein was also made possible by the inventors' use of longitudinal *in vivo* analyses, rather than of *in vitro* analyses. In particular, the inventors' earlier studies used reporter genes (including bacterial β-galactosidase and chloramphenicol acetyltransferase [CAT]) which are heterologous to the factor IX promoter. In these earlier studies, the factor IX promoter showed only very weak expression activity *in vitro* [Kurachi et al. (1995) J. Biol. Chem. 270:5276-5281]. Use of such heterologous reporter genes made it impossible to reliably and quantitatively perform

longitudinal analyses of transgene expression in animals. The inventors unexpectedly observed that the use of hFIX minigene expression vectors which contained the hFIX promoter and its homologous hFIX gene were capable of producing high level plasma hFIX *in vivo*. This unexpected observation not only solved the problems associated with the use of genes which are heterologous to the hFIX promoter by providing a reliable animal assay system, but also provided multiple unexpected critical insights into the regulatory mechanisms of the hFIX gene, including the determination of nucleotide sequences which regulate the stability and age-related increased expression of the exemplary hFIX gene.

The present invention provides the 32-nucleotide nucleic acid sequence 5'-agccatt cagtcgagga aggatagggt ggtat-3' (SEQ ID NO:1) of AE5' which corresponds to the sequence from 2164 to 2195 of the hFIX gene deposited in GenBank as accession number K02402, and which corresponds to the sequence from -802 to -771 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The present invention also provides the 1273-nucleotide nucleic acid sequence (SEQ ID NO:3) (Figure 13) of AE3' which corresponds to the sequence from 34,383 to 35,655 of GenBank accession number K02402, and which corresponds to the sequence from 31,418 to 32,690 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The terms "age-related regulatory activity" and "age-related activity" when made in reference to a nucleic acid sequence refer to the ability of the nucleic acid sequence to alter in an age-related manner (*e.g.*, increase over a period of time) the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence as compared to the level of transcription into mRNA of the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has age-related regulatory activity. An "age regulatory sequence" is herein used to refer to a nucleic acid sequence which has age-related regulatory activity.

To illustrate, where expression levels of a gene of interest decrease over a period of time, a nucleic acid sequence is said to have age-related regulatory activity if (when operably linked to the gene of interest) it results in (a) a smaller decrease in expression levels of the gene over the same period of time as compared to the decrease in expression levels in the absence of the nucleic acid sequence, (b) relatively constant (*i.e.*, unchanged) expression

levels over the same period of time, or (c) increased expression levels over the same period of time.

The terms "operably linked," "in operable combination," and "in operable order" as used herein refer to the linkage of nucleic acid sequences such that they perform their intended function. For example, operably linking a promoter sequence to a nucleotide sequence of interest refers to linking the promoter sequence and the nucleotide sequence of interest in a manner such that the promoter sequence is capable of directing the transcription of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest. Similarly, operably linking a nucleic acid sequence having age-related regulatory activity to a promoter sequence and to a nucleotide sequence of interest means linking the nucleic acid sequence having age-related regulatory activity, the promoter sequence and the nucleotide sequence of interest in a manner such that the nucleic acid sequence having age-related regulatory activity is capable of altering over a period of time the level of transcription into mRNA of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest.

Methods for determining age-related regulatory activity of a candidate nucleic acid sequence, given the teachings of the present specification, are within the ordinary skill in the art and are exemplified by the methods disclosed herein. For example, a test vector is constructed in which the candidate nucleic acid sequence is linked upstream or downstream of a promoter sequence which is operably linked to a nucleotide sequence of interest (*e.g.*, Example 1). A control vector which is similar to the test vector but which lacks the candidate nucleic acid sequence is also constructed. The test vector and control vector are separately introduced into a host cell. It is preferred that the host cell (*e.g.*, fertilized egg) be capable of generating a transgenic multicellular organism, *e.g.*, a transgenic mouse (*e.g.*, Example 3) and that transgenic multicellular organisms are generated. Longitudinal analyses of the expression of mRNA which is encoded by the nucleotide sequence of interest (*e.g.*, by Northern blot hybridization) over a period of time in, and preferably over the entire life span of, founders and successive generations of the transgenic multicellular organism are carried out (*e.g.*, Example 3). The detection in any tissue of mRNA and/or protein levels which are encoded by the nucleotide sequence of interest and which are greater in transgenic animals harboring the test vector as compared to the mRNA and/or protein levels in transgenic

animals harboring the control vector at least one point in time indicates that the candidate nucleic acid sequence has age-related regulatory activity.

For example, evidence provided herein shows the surprising result that AE5' (SEQ ID NO:1) alone has age-related regulatory activity in that AE5' stabilizes hFIX mRNA whereby hFIX mRNA levels are essentially unchanged at different time points over the entire life span of transgenic animals (Figure 4, A, C and E) as compared to the declining hFIX mRNA levels in transgenic animals which harbor vectors that lack AE5' (Figures 2A and 2E). The age-related regulatory activity of AE5' was observed regardless whether AE5' was placed upstream (Figure 4A) or downstream (Figure 4E) of the promoter sequence in the expression construct.

Furthermore, data provided herein demonstrates the unexpected result that AE3' (SEQ ID NO:3) alone has age-related regulatory activity in that AE3' increases hFIX mRNA at several time points during the life of transgenic animals (Figure 2B) relative to the hFIX mRNA levels at the same time points in transgenic animals harboring vectors that lack AE3' (Figure 2A). AE3' substantially increased the steady state hFIX mRNA levels (Figure 5). This result which was observed *in vivo* was surprising in part because AE3' exhibited weak down regulatory effects on hFIX production *in vitro*. Without limiting the invention to any particular mechanism, these results suggest that AE3' functions by increasing hFIX mRNA stability which directly correlates with an increase in the hFIX protein level in the circulation. Also without intending to limit the invention to any particular theory, it is the inventors' consideration that the age-related regulatory activity of AE3' is due to the s1 structure-forming dinucleotide repeats present in the 3'UTR; the s1 region is the 103 bp sequence (SEQ ID NO:61) from nt 32,141 through nt 32,243 of Figure 8. This consideration is based on the inventors' observation that dinucleotide repeats, such as $(AT)_n$ of the 3' UTR of various genes, can form s1 structures in mRNA, which upon binding specific proteins are known to modulate mRNA stability, mostly to a less stable state [Ross (1995) Microbiol. Rev. 59:423-450].

Importantly, the invention demonstrates the surprising synergistic action of AE5' and AE3' which together result in hFIX mRNA levels which not only are greater at each time point tested over the life span of transgenic animals (Figures 4 B and D) as compared to hFIX mRNA levels in transgenic animals harboring vectors that lack both AE5' and AE3', but also that the profile of increased human FIX mRNA levels over the life span of

transgenic mouse recapitulates the profile of increased mouse FIX mRNA levels as a wild-type mouse ages.

Data presented herein further demonstrate that the age-related regulatory activity of AE5' alone, of AE3' alone, and of the combination of AE5' and AE3' is independent of the level of expression of the transgenes harboring them, sex, generation or zygosity status of the transgenic animals.

The present invention is not limited to SEQ ID NOs:1 and 3 but specifically contemplates portions thereof. As used herein the term "portion" when made in reference to a nucleic acid sequence refers to a fragment of that sequence. The fragment may range in size from five (5) contiguous nucleotide residues to the entire nucleic acid sequence minus one nucleic acid residue. Thus, a nucleic acid sequence comprising "at least a portion of" a nucleotide sequence comprises from five (5) contiguous nucleotide residues of the nucleotide sequence to the entire nucleotide sequence.

In a preferred embodiment, portions of SEQ ID NO:1 contemplated to be within the scope of the invention include, but are not limited to, the 7-nucleotide nucleic acid sequence of the polyomavirus enhance activator 3 (PEA-3) (5'-GAGGAAG-3') (SEQ ID NO:2) which corresponds to the sequence from 2176 to 2182 of GenBank accession number K02402, and which corresponds to the sequence from -790 to -784 of GenBank accession number K02402 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30. A nucleotide sequence [5'-CAGGAAG-3' (SEQ ID NO:40)] which is homologous to the invention's PEA-3 nucleotide sequence was initially reported in the art as a polyoma virus enhancer, and was reported to be involved in the regulation of expression of various genes (*e.g.*, collagen gene and *c-fos*) in several tissues [Martin *et al.* (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin *et al.* (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre *et al.*(1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. However, the PEA-3 protein sequence [or PEA-3 related protein(s)] which binds to nucleotide sequences which are homologous to the invention's PEA-3 nucleotide sequence has not been reported to be either liver-specific or enriched in the liver.

Other portions of SEQ ID NO:1 included within the scope of the invention include, for example, SEQ ID NO:33 [5'-tcgaggaagga-3'], SEQ ID NO:34 [5'-agtcgaggaaggata-3'], SEQ ID NO:35 [5'-tcagtcgaggaaggatagg-3'], SEQ ID NO:36 [5'-attcagtcgaggaaggatagggt-3'], SEQ ID NO:37 [5'-ccattcagtcgaggaaggatagggtgg-3'], and

SEQ ID NO:38 [5'-gccattcagtcgaggaaggatagggtggta-3'], all of which include the PEA-3 nucleotide sequence.

In a preferred embodiment, portions of SEQ ID NO:3 contemplated to be within the scope of the invention include, but are not limited to, SEQ ID NO:51 [5'-TTATTTTATATATATAATATATATATAAAATA-3'], SEQ ID NO:52 [5'-TATAATATA-3'], SEQ ID NO:53 [5'-CAATATAAATATATAG-3'], SEQ ID NO:54 [5'-TGTGTGTGTATGCGTGTGTGTAGACACACACGCATACACACATA-3'], the combination of SEQ ID NOs:51 and 52, *i.e.*, SEQ ID NO:55 [5'-TTATTTTATATATATAATATATATATAAAATATATAATATA-3'], the combination of SEQ ID NOs:52 and 53, *i.e.*, SEQ ID NO:56 [5'-TATAATATACAATATAAATATATAG-3'], the combination of SEQ ID NOs:53 and 54, *i.e.*, SEQ ID NO:57 [5'-CAATATAAATATATAGTGTGTGTGTATGCGTGTGTGTAGACACACACGCATACACACATA-3'], the combination of SEQ ID NOs:51, 52, 53, and 54, *i.e.*, SEQ ID NO:58 [5'-TTATTTTATATATATAATATATATATAAAATATATAATATACAATATAAATATATAGTGTGTGTGTATGCGTGTGTGTAGACACACACGCATACACACATA-3'], the 723 bp sequence (SEQ ID NO:59) from nt 31,418 through nt 32,140 of Figure 8, the 447 bp sequence (SEQ ID NO:60) from nt 32,244 through nt 32,690 of Figure 8, and the 103 bp sequence (SEQ ID NO:61) (*i.e.*, the s1 region of the 3' UTR) from nt 32,141 through nt 32,243 of Figure 8.

The nucleotide sequence of portions of SEQ ID NOs:1 and 3 which exhibit age-related regulatory activity may be determined using methods known in the art, *e.g.*, using deletion constructs (*e.g.*, see Yang *et al.* (1998) J. Biol. Chem. 273:891-897). Briefly, several expression plasmids are constructed to contain a reporter gene under the control of a promoter and of different candidate nucleotide sequences which are obtained either by restriction enzyme deletion of internal sequences in SEQ ID NOs:1 and 3, restriction enzyme truncation of sequences at the 5' and/or 3' end of SEQ ID NOs:1 and 3, by the introduction of single nucleic acid base changes by PCR into SEQ ID NOs:1 and 3, or by chemical synthesis. The gene-related regulatory activity of the different constructs is determined as described *supra* in order to determine whether the candidate nucleotide sequence exhibits age-related regulatory activity.

The sequences of the present invention are not limited to SEQ ID NOs:1 and 3 and portions thereof, but also include homologs of SEQ ID NOs:1 and 3, and homologs of portions thereof. Homologs of SEQ ID NOs:1 and 3, and of portions thereof, include, but

are not limited to, nucleotide sequences having deletions, insertions or substitutions of different nucleotides or nucleotide analogs as compared to SEQ ID NOs:1 and 3, and of portions thereof, respectively. Such homologs may be produced using methods well known in the art.

A "homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1. Homologs of SEQ ID NO:1 are exemplified, but not limited to, SEQ ID NO:66 (5'-acccatt cagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the G at nt 2,165 is replaced with a C; SEQ ID NO:67 (5'-agccatt gagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the C at nt 2,171 is replaced with a G; SEQ ID NO:68 (5'-agccatt cagacgagga aggatagggt ggtat-3') is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the T at nt 2,174 is replaced with a A; SEQ ID NO:69 (5'-agccatt cagtcgagga aggatagggt ggttt-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the A at nt 2,194 is replaced with a T; SEQ ID NO:70 (5'-agccatt cagtcgagga tcccaagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that AGGGT beginning at nt 2,186 is replaced with TCCCA; SEQ ID NO:71 (5'-agccatt cagtcgagga aggatagggcctaat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that TGGT beginnining at nt 2,190 is replaced with CCTA; SEQ ID NO:72 (5'-agaccatt cagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a A is inserted after nt 2,165; SEQ ID NO:73 (5'-agccatt cagtcgagga aggatagcggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a C is inserted after nt 2,187; SEQ ID NO:74 (5'-agccatt cagtcgagga aggataat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that GGGTGGT beginning at nt 12,187 is deleted; and SEQ ID NO:75 (5'-agccatt cgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that CAGT beginning at nt 2,171 is deleted.

A "homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2. Homologs of SEQ ID NO:2 include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40),

CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

A "homolog" of SEQ ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3. Homologs of SEQ ID NO:3 are exemplified, but not limited to, SEQ ID NOs:76-83 shown in Figure 13. Specifically, SEQ ID NO:76 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the C at nt 34,390 has been replaced with a G. SEQ ID NO:77 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the T at nt 34,649 has been replaced with a A. SEQ ID NO:78 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the GC beginning at nt 34,959 has been replaced with a CG. SEQ ID NO:79 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the CATG beginning at nt 35,501 has been replaced with a GTAC. SEQ ID NO:80 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TT is inserted after the A at nt 34,681. SEQ ID NO:81 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TGC is inserted after the C at nt 35,581. SEQ ID NO:82 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that A at nt 35,636 is deleted. SEQ ID NO:83 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the G at nt 34,383 is deleted.

A "homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59.

A "homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60.

A "homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61.

Homologs of a portion of SEQ ID NO:1 are exemplified by homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2), which include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40), CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

The present invention also contemplates functioning or functional homologs of SEQ ID NO:1, of portions of SEQ ID NO:1 (*e.g.*, functional portions of SEQ ID NOs:2, and 33-38), of SEQ ID NO:3, and of portions of SEQ ID NO:3 (*e.g.*, functional portions of SEQ ID NOs:51-61).

A "functional homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and which has age-related regulatory activity. Alternatively, a functional homolog of SEQ ID NO:1 is a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and having liver-specific activity.

A "functional homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and which has age-related regulatory activity. Alternatively, a functional homolog of SEQ ID NO:2 is a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and having liver-specific activity.

A "functional homolog" of SEQ ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61, and which has age-related regulatory activity.

The present invention is not limited to sense molecules of SEQ ID NOs:1 and 3 but contemplates within its scope antisense molecules comprising a nucleic acid sequence complementary to at least a portion (*e.g.*, a portion greater than 10 nucleotide bases in length and more preferably greater than 100 nucleotide bases in length) of the nucleotide sequence of SEQ ID NOs:1 and 3. These antisense molecules find use in, for example, reducing or

preventing expression of a gene (*e.g.* hFIX) whose expression is regulated by SEQ ID NOs:1 and 3.

The invention further provides the nucleotide sequence AE3'' which is a preferred portion of AE3'. AE3'' [5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; (SEQ ID NO:93)] is the 154-nucleotide nucleic acid sequence from 35,075 to 35,228 of GenBank accession number K02402, which corresponds to the sequence from 32,110 to 32,263 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30. AE3'' contains a 102-bp stem-loop forming sequence (SEQ ID NO:91).

Data presented herein demonstrates the universality of the regulatory function of portions of AE3' in that AE3'' has been successfully used to regulate expression of the heterologous hPC gene in an age-related manner. In particular, transgenic animals containing the -1462hPCm1 construct expressed age-stable levels of human protein C, *i.e.*, expressed relatively constant levels of human protein C at different time points during the life span of the transgenic animals (Example 12, Figure 17A). In direct contrast, the presence of AE5' and AE3'' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of the AE3'' portion of AE3' in regulating expression of operably linked genes in an age-related manner.

The invention also contemplates portions of AE3''. These portions include, but are not limited to, SEQ ID NOs:94-144 wherein SEQ ID NO:94 is 5'-tgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:95 is 5'-gggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:96 is 5'-ggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:97 is 5'-gg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:98 is 5'-g gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:99 is 5'-

gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:100 is 5'-aaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:101 is 5'-aaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:102 is 5'-aagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:103 is 5'-agtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:104 is 5'-gtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:105 is 5'-tttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:106 is 5'-ttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:107 is 5'-tc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:108 is 5'-c tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:109 is 5'-tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:110 is 5'-ttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:111 is 5'-tcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:112 is 5'-cagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:113 is 5'-agagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:114 is 5'-gagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:115 is 5'-agag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata

cacacatata atggaagcaa taagccat-3'; SEQ ID NO:116 is 5'-gag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:117 is 5'-ag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:118 is 5'-g ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:119 is 5'-ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:120 is 5'-taagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:121 is 5'-aagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:122 is 5'-agttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:123 is 5'-gttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:124 is 5'-ttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:125 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcca-3'; SEQ ID NO:126 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcc-3'; SEQ ID NO:127 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagc-3'; SEQ ID NO:128 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taag-3'; SEQ ID NO:129 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taa-3'; SEQ ID NO:130 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa ta-3'; SEQ ID NO:131 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa t-3'; SEQ ID NO:132 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata

atggaagcaa -3'; SEQ ID NO:133 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagca-3'; SEQ ID NO:134 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagc-3'; SEQ ID NO:135 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaag-3'; SEQ ID NO:136 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaa-3'; SEQ ID NO:137 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgga-3'; SEQ ID NO:138 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgg-3'; SEQ ID NO:139 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atg-3'; SEQ ID NO:140 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata at-3'; SEQ ID NO:141 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata a-3'; SEQ ID NO:142 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata-3'; SEQ ID NO:143 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatat-3'; and SEQ ID NO:144 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacata-3'.

In a particularly preferred embodiment, the portion of AE3' and of AE3'' is the 102-nucleotide nucleic acid stem-loop forming sequence 5'-tatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacata-3' (SEQ ID NO:91). The stem-loop sequence is located from nt 32,142 to 32,243 of Figure 8, which corresponds to the sequence from nt 35,107 to 35,208 of GenBank accession number K02402. It is the inventors' belief that the 102-nt stem-loop forming sequence exhibits the age-regulatory functions, described herein, of the AE3' and AE3'' sequences.

The nucleotide sequence of SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be synthesized by synthetic chemistry techniques which are commercially available and well known in the art [see Caruthers MH *et al.*, (1980) Nuc. Acids Res. Symp. Ser. 215-223; Horn T. *et al.*, (1980) Nuc. Acids Res. Symp. Ser. 225-232]. Additionally, fragments of SEQ ID NOs:1 and 3 can be made by treatment of SEQ ID NOs:1 and 3 with restriction enzymes followed by purification of the fragments by gel electrophoresis. Alternatively, sequences may also be produced using the polymerase chain reaction (PCR) as described by Mullis [U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference]. SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be ligated to each other or to heterologous nucleic acid sequences using methods well known in the art.

The nucleotide sequence of synthesized sequences may be confirmed using commercially available kits as well as using methods well known in the art which utilize enzymes such as the Klenow fragment of DNA polymerase I, Sequenase®, *Taq* DNA polymerase, or thermostable T7 polymerase. Capillary electrophoresis may also be used to analyze the size and confirm the nucleotide sequence of the products of nucleic acid synthesis, restriction enzyme digestion or PCR amplification.

It is readily appreciated by those in the art that the sequences of the present invention may be used in a variety of ways. For example, the nucleic acid sequences of the invention and portions thereof can be used as probes for the detection and isolation of functional homologs of AE5' and AE3', amplification of homologous nucleotide sequences, age-related and/or liver-specific expression of a nucleotide sequence of interest in an animal, gene therapy, and reducing factor IX levels in an animal.

ii. Regulatory Nucleic Acid Sequences From The hPC Gene

The invention provides regulatory nucleic acid sequences which are derived from the hPC gene. The presence of these sequences and their surprising properties was fortuitously discovered by the inventors during their investigation of the universality of the function of regulatory sequences from the hFIX gene. In particular, when using what they believed would be "control" constructs which contained different portions of the sequence upstream of nt +1 in the hPC gene, the inventors discovered that whereas transgenic animals containing the -1462hPCm1 construct exhibited relatively constant and relatively high levels (from

about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCm1 construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) (Figure 17B) of human protein C which declined at a precipitous rate over time (Example 12). Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCm1 construct. These results demonstrated to the inventors that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene exhibits age-related regulatory activity (as evidenced by age-stable expression over time) and regulatory activity (as evidenced by the relatively higher levels of expression as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.

The term "regulatory activity" when made herein in reference to a nucleic acid sequence refers to the ability of the nucleic acid sequence to alter the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence, as compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. In contrast to "age-related regulatory activity," the term "regulatory activity" relates to alteration in the level of transcription or protein synthesis at a single time point (rather than over a period of time) in the life cycle of a cell, tissue, or organism. In a preferred embodiment, the level of transcription into mRNA and/or synthesis of polypeptide is increased from 2-fold to at least 10,000-fold, preferably from 2-fold to 10,000 fold, more preferably from 2-fold to 1,000-fold, and most preferably, from 2-fold to 500-fold, when compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. Data presented herein demonstrates that the inclusion of the human protein C (hPC) nucleotide sequence from nt -1462 to nt -83 resulted in expression of from about 100 to about 3000 ng/ml of human protein C (Figure 17A) as compared to expression of from about 5 to about 40 ng/ml of human protein C (Figure 17B) in transgenic mice at one-month of age. The increased level of expression was even more dramatic when the transgenic animals were 5 months-old; while the levels of hPC were undetectable in the

absence of the sequence from nt -1462 to nt -83, they were from about 100 to about 3000 ng/ml in the presence of this sequence.

In one embodiment, the invention provides the nucleic acid sequence (SEQ ID NO:85) from -1462 to +1 of the hPC gene (Figure 14; 5'-GAATTCTGTA AGCATTTCCT ATGTGTACCT GCCCCTGGGC AAGGTGGGCC TGACTTGTTA GAGTGTTAGA GTTTTACCCT GTTCCTCTAG GAGGGCCTGG TACCACCACA GCCCAGCATG GTGTGGTGCC TCAGCAGGAG GCATCTGGTT ACAATCAACA CAAGCTGTTC CAGCCAATTT AAAGAAACTT CAGGAGGAAT AGGGTTTTAG GAGGGCATGG GGACCCTCCT GCACCCGAAG CCAGGATGTG CCACCAATCA TAAGGAGGCA GGGGCCTCCT TCCGCTGCTC CCTGGGACTC TCTAGGTGTC CGTGGCCTCA GCCCCCCTCT GCACACCTGC ATCTTCCTTC TCATCAGCTT CCTCTGCTTT AAGCGTAAAC ATGGATGCCC AGGACCTGGC CTCAATCTTC CGAGTCTGGT ACTTATGGTG TACTGACAGT GTGAGACCCT ACTCCTCTGA TCAATCCCCT GGGTTGGTGA CTTCCCTGTG CAATCAATGG AAGCCAGCGA GGCAGGGTCA CATGCCCCGT TTAGAGGTGC AGACTTGGAG AAGGAACGTG GGCAAGTCTT CCCAGGAACA GGTAGGGCAG GGAGGAAAGG GGGGCATCTC TGGTGCAGCC CGGTTCGGAG CAGGAAGACG CTTAATAAAT GCTGATAGAC TGCAGGACAC AGGCAAAGGT GCTGAGCTGG ACCCTTTATT TCTGCCCTTC TCCCTTCTGG CACCCCGGCC AGGAAATTGC TGCAGCCTTT CTGGAATCCC GTTCATTTTT CTTACTGGTC CACAAAAGGG GCCAAATGGA AGCAGCAAGA CCTGAGTTCA AATTAAATCT GCCAACTACC AGCTCAGTGA ATCTGGGCGA GTAACACAAA ACTTGAGTGT CCTTACCTGA AAAATAGAGG TTAGAGGGAT GCTATGTGCC ATTGTGTGTG TGTGTTGGGG GTGGGGATTG GGGGTGATTT GTGAGCAATT GGAGGTGAGG GTGGAGCCCA GTGCCCAGCA CCTATGCACT GGGGACCCAA AAAGGAGCAT CTTCTCATGA TTTTATGTAT CAGAAATTGG GATGGCATGT CATTGGGACA GCGTCTTTTT TCTTGTATGG TGGCACATAA ATACATGTGT CTTATAATTA ATGGTATTTT AGATTTGACG AAATATGGAA TATTACCTGT TGTGCTGATC TTGGGCAAAC TATAATATCT CTGGGCAAAA ATGTCCCCAT CTGAAAAACA GGGACAACGT TCCTCCCTCA GCCAGCCACT ATGGGGCTAA AATGAGACCA CATCTGTCAA GGGTTTTGCC CTCACCTCCC TCCCTGCTGG ATGGCATCCT TGGTAGGCAG AGGTGGGCTT CGGGCAGAAC AAGCCGTGCT GAGCTAGGAC CAGGAGTGCT AGTGCCACTG TTTGTCTATG GAGAGGGAGG

CCTCAGTGCT GAGGGCCAAG CAAATATTTG TGGTTATGGA TTA-3'). This sequence was successfully used to express hPC in transgenic mice in an age-related manner (*i.e.*, at relatively constant levels over time; Figure 17A), and at relatively high levels (*e.g.*, as compared to expression levels in the presence of sequences from nt -1462 to nt -82; Figure 17B).

The invention also provides the nucleotide sequence (SEQ ID NO:92) from -1462 to -83 of the hPC gene (Figure 14). The finding that SEQ ID NO:92 exhibited age-related regulatory activity and regulatory activity as discussed in Example 11, *infra*, was surprising because it was contrary to the results which had previously been reported by Miao et al. [Miao et al. (1996) J. Biol. Chem. 16:9587-9594] when using a heterologous reporter gene, chloramphenicol acetyltranferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences. In particular, Miao et al. found that construct pPC-1528 which contained nucleotides -1462 to +1 resulted in substantially reduced CAT *in vitro* activity as compared to construct pPC-82-66 which contained nucleotides -82 to +1. From this, Miao et al: concluded that there is an element with silencer activity in the region between -1462 and -82. In contrast, data presented herein demonstrates that nucleotides -1462 to +1 and -82 to +1 resulted in similar *in vitro* activities of hPC (Example 11, Figure 16). Indeed, the inventors' data disclosed in relation to hPC's age-related regulatory sequences which are upstream of the hPC coding sequences when used to regulate expression of hPC is consistent with the inventor's observation (Figure 1) in relation to hFIX's age-related regulatory AE5' sequences when used to regulate expression of hFIX. In particular, the inventors have observed that hFIX's AE3' and AE3'' sequences resulted in moderate suppression in *in vitro* transient expression assays when using hFIX and hPC, respectively (Figures 1 and 16).

Further contemplated to be within the scope of this invention are portions of the SEQ ID NOS:85 and 92. In a particularly preferred embodiment, these portions contain one or both of the first PEA-3 element (SEQ ID NO:89) [7-bp long; 5'-GAGGAAA-3', from -871 to -865 of the hPC gene of Figure 14] and the second PEA-3 element (SEQ ID NO:90) [7-bp long; 5'-CAGGAAG-3', from -832 to -826 of the hPC gene of Figure 14].

Exemplary portions of SEQ ID NOs:85 and 92 which contain both the first and second PEA-3 elements include, but are not limited to, the nucleotide sequence (SEQ ID NO:88) from -1462 to -802 of the hPC gene, which is embodied in plasmid -849hPCm1

(Figure 15, Example 10). Further examples include the nucleotide sequence (SEQ ID NO:145) from -1462 to -804, (SEQ ID NO:146) from -1462 to -805, (SEQ ID NO:147) from -1462 to -806, (SEQ ID NO:148) from -1462 to -807, (SEQ ID NO:149) from -1462 to -808, (SEQ ID NO:150) from -1462 to -809, (SEQ ID NO:151) from -1462 to -810, (SEQ ID NO:152) from -1462 to -811, (SEQ ID NO:153) from -1462 to -812, (SEQ ID NO:154) from -1462 to -813, (SEQ ID NO:155) from -1462 to -814, (SEQ ID NO:156) from -1462 to -815, (SEQ ID NO:157) from -1462 to -816, (SEQ ID NO:158) from -1462 to -817, (SEQ ID NO:159) from -1462 to -818, (SEQ ID NO:160) from -1462 to -819, (SEQ ID NO:161) from -1462 to -820, (SEQ ID NO:162) from -1462 to -821, (SEQ ID NO:163) from -1462 to -822, (SEQ ID NO:164) from -1462 to -823, (SEQ ID NO:165) from -1462 to -824, (SEQ ID NO:166) from -1462 to -825, (SEQ ID NO:167) from -1462 to -826, (SEQ ID NO:168) from -1452 to -803, (SEQ ID NO:169) from -1442 to -803, (SEQ ID NO:170) from -1412 to -803, (SEQ ID NO:171) from -1102 to -803, (SEQ ID NO:172) from -902 to -803, and (SEQ ID NO:173) from -873 to -803.

Portions of SEQ ID NOs:85 and 92 which contain only the first PEA-3 element include the nucleotide sequence (SEQ ID NO:87) from -1462 to -849 of the hPC gene, which is embodied in plasmid -802hPCm1 (Figure 15, Example 10). Additional exemplary portions include, but are not limited to the nucleotide sequence (SEQ ID NO:174) from -1462 to -850, (SEQ ID NO:175) from -1462 to -851, (SEQ ID NO:176) from -1462 to -852, (SEQ ID NO:177) from -1462 to -853, (SEQ ID NO:178) from -1462 to -854, (SEQ ID NO:179) from -1462 to -855, (SEQ ID NO:180) from -1462 to -856, (SEQ ID NO:181) from -1462 to -857, (SEQ ID NO:182) from -1462 to -858, (SEQ ID NO:183) from -1462 to -859, (SEQ ID NO:184) from -1462 to -860, (SEQ ID NO:185) from -1462 to -861, (SEQ ID NO:186) from -1462 to -862, (SEQ ID NO:187) from -1462 to -863, (SEQ ID NO:188) from -1462 to -864, (SEQ ID NO:189) from -1462 to -865, (SEQ ID NO:190) from -1362 to -865, (SEQ ID NO:191) from -1262 to -865, (SEQ ID NO:192) from -1162 to -865, (SEQ ID NO:193) from -1062 to -865, (SEQ ID NO:194) from -962 to -865, and (SEQ ID NO:195) from -872 to -865.

Examples of portions of SEQ ID NOs:85 and 92 which contain only the second PEA-3 element include, but are not limited to, the nucleotide sequence (SEQ ID NO:196) from -863 to -83, (SEQ ID NO:197) from -853 to -83, (SEQ ID NO:198) from -843 to -83, (SEQ ID NO:199) from -833 to -83, (SEQ ID NO:200) from -832 to -83, (SEQ ID NO:201) from

-863 to -183, (SEQ ID NO:202) from -863 to -283, (SEQ ID NO:203) from -863 to -383, (SEQ ID NO:204) from -863 to -483, (SEQ ID NO:205) from -863 to -583, (SEQ ID NO:206) from -863 to -683, (SEQ ID NO:207) from -863 to -783, and (SEQ ID NO:208) from -863 to -826.

In a particularly preferred embodiment, the portion of SEQ ID NOs:85 and 92 is selected from the first PEA-3 element (SEQ ID NO:89) and the second PEA-3 element (SEQ ID NO:90). It is the inventor's view that the first and/or second PEA-3 elements within SEQ ID NOs:85 and 92 are responsible for the observed age-related regulatory activity and regulatory activity of SEQ ID NOs:85 and 92.

B. Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

The invention provided herein is not limited to SEQ ID NO:1, 3, 85 and 92, homologs and portions thereof having age-related regulatory activity, but includes sequences having no age-related regulatory activity (*i.e.*, non-functional homologs and non-functional portions of homologs). The use of such sequences may be desirable, for example, where a portion of SEQ ID NOs:1, 3, 85, and 92 is used as a probe to detect the presence of SEQ ID NOs:1, 3, 85 and 92, respectively, or of portions thereof in a sample.

As used herein, the term "probe" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to a nucleotide sequence of interest. A probe may be single-stranded or double-stranded. It is contemplated that any probe used in the present invention will be labelled with any "reporter molecule," so that it is detectable in any detection system including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, calorimetric, gravimetric, magnetic, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The probes provided herein are useful in the detection, identification and isolation of, for example, sequences such as those listed as SEQ ID NOs:1, 3, 85 and 92 as well as of homologs thereof. Preferred probes are of sufficient length (*e.g.*, from about 9 nucleotides to about 20 nucleotides or more in length) such that high stringency hybridization may be

employed. In one embodiment, probes from 20 to 50 nucleotide bases in length are employed.

C. Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

The invention provided herein is not limited to SEQ ID NOs:1 and 3, homologs and portions thereof having age-related regulatory activity, but includes sequences having no age-related regulatory activity. This may be desirable, for example, where a portion of the nucleic acid sequences set forth as SEQ ID NOs:1 and 3 is used as a primer for the amplification of nucleic acid sequences by, for example, polymerase chain reactions (PCR) or reverse transcription-polymerase chain reactions (RT-PCR). The term "amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art [Dieffenbach CW and GS Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview NY]. As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis disclosed in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter

"PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are the to be "PCR amplified."

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; and/or incorporation of ^{32}P-labeled deoxyribonucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any nucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (*e.g.*, genes) for the construction of targeting vectors, transgenes, *etc.*

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long (*e.g.*, from about 9 nucleotides to about 20 nucleotides or more in length) to prime the synthesis of extension products in the presence of the inducing agent. Suitable lengths of the primers may be empirically determined and depend on factors such as temperature, source of primer and the use of the method. In one embodiment, the present invention employs primers from 20 to 50 nucleotide bases in length.

The primers contemplated by the invention are useful in, for example, identifying sequences which are homologous to AE5', AE3', and regulatory sequences derived from hPC, in mammals, yeast, bacteria, and in other organisms.

D. Methods For Regulating Gene Expression

The present invention provides methods for regulating expression of a nucleotide sequence of interest over a period of time in a cell or multicellular organism. Specifically, gene expression is preferably regulated in a multicellular organism. In one embodiment, expression of a nucleotide sequence of interest is stabilized such that the level of mRNA and/or protein encoded by the nucleotide sequence of interest remains relatively unchanged at different times during the life of the organism. In an alternative embodiment, expression of a nucleotide sequence of interest is increased. Increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest at a given time point is greater than the level of mRNA and/or protein, respectively, at an earlier time point during the life of the organism or cell. Alternatively, increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest is greater than the level of mRNA and/or protein, respectively, at the same time point in the life of the organism or cell as compared to the level of mRNA and/or protein when expressed in the absence of the sequences of the invention.

In one embodiment, regulating expression of a nucleotide sequence of interest over a period of time is accomplished by introducing into a host cell a vector that contains a nucleotide sequence of interest operably linked to a promoter sequence and to sequences provided herein which have age-related regulatory activity. The transfected host cell is allowed to develop into a transgenic animal in which the nucleotide sequence of interest is expressed in at least one tissue. These steps are further described below for specific embodiments.

1. Expression Constructs

In one embodiment of the methods of the invention for regulating expression of a nucleotide sequence of interest in an age-related manner and/or to liver tissue, a vector is constructed in which the nucleic acid sequences of the invention (*e.g.*, AE5' alone, AE3' alone, or a combination of AE5' and AE3') are operably linked to a promoter sequence and to a nucleotide sequence of interest. In one embodiment, the nucleotide sequence of interest is the coding region of the hFIX gene (Example 1). In another embodiment the nucleotide sequence of interest is the coding region of the protein C gene (Example 7).

The invention is not limited to coding sequences of the hFIX gene or protein C gene. Rather, any nucleotide acid sequence whose expression is desired to be regulated by sequences provided herein are contemplated to be within the scope of this invention. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes which encode a protein [*e.g.*, reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factor genes, activator protein 1 gene, activator protein 2 gene, Sp1 gene, *etc.*]. In one preferred embodiment, the structural gene is the human α1-antitrypsin gene (Figure 10) (SEQ ID NO:42) which encodes a plasma proteinase inhibitor used for treating emphysema. In another preferred embodiment, the structural gene is one encoding the human antithrombin III (Figure 11) (SEQ ID NO:43) which is a plasma proteinase inhibitor for activated blood coagulation factors and whose activity is increased by heparin. In yet another preferred embodiment, the structural gene is the gene encoding the PEA-3 protein (Figure 9) (SEQ ID NO:47) and/or its related protein, which has been shown to bind specifically to homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2) disclosed herein.

The invention is not limited to using a single nucleotide sequence of interest in operable combination with the sequences of the invention. Rather, a plurality (*i.e.*, more than one) of nucleotide sequences of interest may be ligated in tandem such that their expression is regulated by the regulatory sequences of the invention. A plurality of coding sequences may be desirable, for example, where it is useful to express a transcription product of more than one gene to permit interaction of these transcriptional products. Alternatively, a plurality of coding sequences may be desirable where one of the gene sequences is a reporter gene sequence. For example, it may be advantageous to place a coding sequence of a reporter gene in tandem with the coding sequence of a gene of interest such that expression of the coding region of both the reporter gene and the gene of interest is regulated by the sequences of the invention. Expression of the reporter gene usually correlates with expression of the gene of interest. Examples of reporter gene sequences include the sequences encoding the enzymes β-galactosidase and luciferase. Fusion genes may also be desirable to facilitate purification of the expressed protein. For example, the heterologous sequence which encodes protein A allows purification of the fusion protein on immobilized immunoglobulin. Other affinity traps are well known in the art and can be utilized to advantage in purifying the expressed fusion protein. For example, pGEX vectors (Promega, Madison WI) may be used to express the polypeptides of interest as a fusion protein with

glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Other fusion polypeptides useful in the purification of proteins of interest are commercially available, including histidine tails (which bind to Ni^{2+}), biotin (which binds to streptavidin), and maltose-binding protein (MBP) (which binds to amylose). Proteins made in such systems may be designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released at will from the heterologous polypeptide moiety to which it is fused.

One of skill in the art would understand that where a plurality of nucleotide sequences of interest is operably linked to sequences of the present invention, the nucleotide sequences of interest may be either contiguous or separated by intervening polynucleotide sequences, so long as the nucleic acid sequences of interest are operably linked to the promoter sequence, and so long as the sequences of the invention are operably linked to the promoter sequence.

While specific preferred embodiments used herein disclose the use of the hFIX promoter and the CMV promoter, it is not intended that the invention be limited to the type or source of the promoter sequence which is operably linked to the sequences of the invention. Any promoter whose activity is desired to be regulated by the sequences provided herein is contemplated to be within the scope of the invention. Exemplary promoters include the tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter (can be isolated from vector plasmid pRc/RSV from Invitrogen), retrovirus LTR promoter (can be isolated from vector plasmid pLXSN from Clontech) SV40 promoter (located between positions +3530 to +3192 in vector plasmid pCR3 from Invitrogen), PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T_7 promoter having the 23-bp sequence (SEQ ID NO:44) 5'-TAATACGACTCACTATAGGGCGA-3', T_3 promoter having the 24-bp sequence (SEQ ID NO:45) 5'-TTATTAACCCTCACTAAAGGGAAG -3', SP6 promoter having the 23-bp sequence (SEQ ID NO:46) 5'-ATTTAGGTGACACTATAGAATAC -3', and K11 promoter. The T_7 promoter, T_3 promoter, SP6 promoter and K11 promoter have been described in U.S Patent No. 5,591,601, the entire contents of which are incorporated by reference.

Nor is the invention intended to be limited to the use of a single promoter. For example, chimeric promoters (*i.e.*, two or more promoters which are derived from at least

one gene) are expressly contemplated to be within the scope of the invention. Such chimeric promoters may be desirable where, for example, chimeric promoters result in increased levels of expression of an operably linked downstream coding sequence. Chimeric promoters are known in the art and include, for example, the double *Tet* promoter [Kistner et al. (1996) Proc. Natl. Acad. Sci. USA 93:10933-10938], and the U1 snRNA promoter-CMV promoter/enhancer [Bartlett et al. (1996) Proc. Natl. Acad. Sci. USA 93:8852-8857].

Expression vectors in which expression of a nucleic acid sequence of interest is regulated by sequences of the invention may be constructed using the teachings of the present invention in conjunction with techniques well known in the art. [Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY; Ausubel *et al.* (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY]. Briefly, the nucleic acid sequence of interest is placed in operable combination with a promoter sequence and sequences of the invention in the presence of transcription and translation regulatory sequences, including initiation signals such as a start codon (*i.e.*, ATG), enhancers, and transcription termination signals. The ATG initiation codon must be in the correct reading frame to ensure translation of the entire heterologous nucleotide sequence. Transcription termination signals are placed downstream of the heterologous nucleic acid sequence and include polyadenylation sequences which are exemplified by, but not limited to, SV40 poly-A sequence, hINV poly-A sequence, or bovine growth hormone poly-A sequence, *etc.*

Other regulatory sequences which may affect RNA stability as well as enhancers (*i.e.*, a sequence which when activated results in an increase in the basal rate of transcription of a gene) and silencers (*i.e.*, a sequence involved in reducing expression of a gene) may also be included. These regulatory sequences may be relatively position-insensitive, *i.e.*, the regulatory element will function correctly even if positioned differently in relation to the heterologous nucleotide sequence in the construct as compared to its position in relation to the corresponding heterologous nucleotide sequence in the genome. For example, an enhancer may be located at different distances from the promoter sequence, in a different orientation, and/or in a different linear order. Thus, an enhancer that is located 3' to a promoter sequence in germline configuration might be located 5' to the promoter sequence in the construct.

2. Host Cells

Host cells are transformed with expression vectors which contain the sequences of the invention in operable combination with a nucleic acid sequence of interest using methods known in the art. The term "transformation" as used herein refers to the introduction of a transgene into a cell. The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence." The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, *etc.*) relative to the naturally-occurring sequence. The terms "heterologous DNA sequence" and "foreign DNA sequence" refer to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification (*e.g.*, a point mutation, the presence of a selectable marker gene, *etc.*) relative to the naturally-occurring gene. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (*e.g.*, proteins which confer drug resistance), *etc.*

Transformation may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, biolistics (*i.e.*, particle bombardment) and the like.

Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one

or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein encoded by the transgene. For example, the activity of β-glucuronidase (GUS) which is encoded by the *uid A* gene may be detected using either a histochemical assay of GUS enzyme activity by staining with X-gluc which gives a blue precipitate in the presence of the GUS enzyme, or a chemiluminescent assay using the GUS-Light kit (Tropix). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

Suitable host cells include bacterial, yeast, plant, insect, and mammalian cells. In one embodiment the host cell is mammalian. In a preferred embodiment, the mammalian host cell is a mouse fertilized egg cell. In an alternative embodiment, the mammalian host cell is a HepG2 cell (ATCC number HB8065), a fibroblast cell (*e.g.*, ATCC number CCL 110), a myoblast cell (*e.g.*, Clonetics, catalog # SkMC), and an endothelial cell (*e.g.*, human umbilical cord endothelial cells; ATCC number CRL 1730).

In one embodiment, the host cell is transformed both with an expression vector which contains the sequences of the invention in operable combination with the nucleic acid sequences of interest, as well as with an expression vector which expresses the PEA-3 protein (Example 6). Such co-transformation may be desirable, for example, where expression of the nucleotide sequence of interest is regulated by AE5' or portions or homologs thereof which contain homologs of the PEA-3 nucleotide sequence to which the PEA-3 protein binds. In one embodiment, expression of the PEA-3 protein is under the control of the LTR promoter of the Moloney murine leukemia virus (MoLV) which is capable of driving expression of operably linked genes in several cell types. Transient

expression assays are suitable for determining the relative promoter activities in expressing desirable PEA-3 protein levels.

Any number of selection systems may be used to recover transfected cells. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M *et al.* (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I *et al.* (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* which confers resistance to methotrexate [Wigler M *et al.*, (1980) Proc Natl Acad Sci 77:3567-70]; *npt*, which confers resistance to the aminoglycosides neomycin and G-418 [Colbere-Garapin F *et al.*, (1981) J. Mol. Biol. 150:1-14] and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine [Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51]. Recently, the use of a reporter gene system which expresses visible markers has gained popularity with such markers as β-glucuronidase and its substrate (GUS), luciferase and its substrate (luciferin), and β-galactosidase and its substrate (X-Gal) being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system [Rhodes CA *et al.* (1995) Methods Mol Biol 55:121-131].

The presence or expression of the reporter gene usually indicates the presence or expression, respectively, of the tandem heterologous nucleic acid sequence as well. However, it is preferred that the presence and expression of the desired heterologous nucleic acid sequence be confirmed. This is accomplished by procedures known in the art which include DNA-DNA or DNA-RNA hybridization or amplification using probes, or fragments of the heterologous nucleic acid sequence. For example, Fluorescent In Situ Hybridization (FISH) can be used to detect the heterologous nucleic acid sequence in cells. Several guides to FISH techniques are available, *e.g.*, Gall *et al.* Meth. Enzymol. 21:470-480 (1981); Angerer *et al.*, in "Genetic Engineering: Principles and Methods," Setlow & Hollaender, Eds. Vol. 7 pp. 43-65, Plenum Press, New York (1985). Alternatively, DNA or RNA can be isolated from cells for detection of the transgene by Southern or Northern hybridization or by amplification based assays. Nucleic acid amplification based assays involve the use of

oligonucleotides or oligomers based on the nucleotide sequence of interest in order to detect cells and tissues which contain the DNA or RNA encoding the transgene of interest. As used herein, the terms "oligonucleotides" and "oligomers" refer to a nucleic acid sequence of at least about five (5) contiguous nucleotide residues and as many as about sixty (60) nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer. Standard PCR methods useful in the present invention are described by Innis *et al.* (Eds.), "PCR Protocols: A Guide to Methods and Applications," Academic Press, San Diego (1990).

Yet another alternative for the detection of heterologous nucleic acid sequences includes detecting the polypeptide product of transcription of the heterologous nucleotide sequence. A variety of protocols which employ polyclonal or monoclonal antibodies specific for the protein product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A competitive binding assay may also be used. Alternatively, a two-site, monoclonal-based immunoassay which utilizes monoclonal antibodies that are reactive to two non-interfering epitopes on the protein of interest may be employed. These and other assays are described in, among other places, Hampton R *et al.* (1990), *Serological Methods a Laboratory Manual*, APS Press, St Paul MN), and Maddox DE *et al.* (1983), J. Exp. Med. 158:1211.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence of interest, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides. A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like.

Host cells transformed with expression vectors containing the sequences provided herein are useful for age-related expression of recombinant proteins of interest. Host cells transformed with expression vectors containing the invention's sequences may be part of a tissue or organ of a living animal. A "living animal" as used herein refers to any multicellular animal (*e.g.*, humans, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.) into whose cells the sequences provided herein may be introduced. Where the host cells (*e.g.*, fertilized egg cells) are capable of generating a multicellular organism, these cells when transformed with expression vectors containing the sequences of the invention are useful in generating transgenic animals which exhibit age-related and/or liver-specific expression of nucleotide sequences of interest.

3. Transgenic Animals

The present invention provides transgenic non-human animals which express a nucleotide sequence of interest in an age-related manner. These animals provide useful models for diseases (*e.g.*, thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) which are associated with increasing age, as well a for screening candidate therapeutic agents against such diseases. These transgenic animals are also useful in studies of normal phenomena, such as ageing, gene regulation, *etc.* In one embodiment, the invention discloses transgenic mice which express in an age-related manner the exemplary hFIX coding sequence under the control of AE5' and/or AE3' (Example 3).

The term "age-related manner" when made in reference to the expression of a nucleotide sequence of interest is a relative term which refers to an increase over a period of time in the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to a promoter and to a nucleic acid sequence which has age-related regulatory activity, as compared to the quantity of mRNA and/or protein, respectively, encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to the promoter in the absence of the nucleic acid sequence which has age-related regulatory activity. Thus, the term "age-related" when made in reference to expression of a nucleotide sequence of interest by a transgenic

animal means that the transgenic animal expresses the nucleotide sequence of interest in an age-related manner.

For example, in one embodiment, the invention demonstrates that hFIX is expressed in an age-related manner in transgenic mice which harbor a transgene (-416FIXm1/1.4) (Figure 2B) which contains hFIX under the control of the hFIX promoter and the regulatory control of AE3' as compared to expression of hFIX in transgenic mice which harbor a transgene (-416FIXm1) (Figure 2A) in which hFIX is under the control of the hFIX promoter in the absence of AE3'. While transgenic mice harboring the -416FIXm1/1.4 construct showed decreasing hFIX activity levels over a period of time (*e.g.*, from 1 to 9 months of age), this decrease was less than the decrease in hFIX activity levels which was observed in transgenic mice harboring the -416FIXm1 construct over the same period of time.

In another embodiment, the invention discloses that hFIX is expressed in an age-related manner in transgenic mice which harbor transgenes (-802FIXm1, -2231FIXm1, and -416FIXm1/AE5') (Figures 4A, C and E) each of which contains hFIX under the control of the hFIX promoter and the regulatory control of AE5' as compared to expression of hFIX in transgenic mice which harbor a transgenes (-416FIXm1 and -770FIXm1) (Figures 2A and E) in which hFIX is under the control of the hFIX promoter in the absence of AE5'. Transgenic mice harboring each of the -802FIXm1, -2231FIXm1, and -416FIXm1/AE5' constructs showed relatively unchanged hFIX activity levels over a period of time (*e.g.*, from 1 to 7 months of age) while transgenic mice harboring either the -416FIXm1 or -770FIXm1 construct showed decreasing hFIX activity levels over the same time period.

In an additional embodiment, the invention shows that hFIX is expressed in an age-related manner in transgenic mice which harbor transgenes (-802FIXm1/1.4 and -2231FIXm1/1.4) (Figures 4B and D) each of which contains hFIX under the control of the hFIX promoter and the regulatory control of both AE3' and AE5' as compared to expression of hFIX in transgenic mice which harbor a transgene (-770FIXm1) (Figure 2E) in which hFIX is under the control of the hFIX promoter in the absence of both AE3' and AE5'. Transgenic mice harboring either the -802FIXm1/1.4 or the -2231FIXm1/1.4 construct showed increasing levels of hFIX activity over a period of time (*e.g.*, 1 to 3 months of age) as compared to decreasing hFIX activity levels over the same period of time in transgenic mice harboring the -770FIXm1 construct.

The present invention also provides transgenic non-human animals which express a nucleotide sequence of interest in a liver-specific manner. These animals are useful for targeting expression of a nucleotide sequence of interest to the liver. Examples of nucleotide sequences of interest are those which encode blood coagulation factors (*e.g.*, factor VIII, factor VII, factor X and prothrombin) whose deficiency is known to play a role in abnormal bleeding disorders. Other examples of nucleotide sequences of interest include those which encode blood coagulation regulators and/or inhibitors (*e.g.*, protein C, antithrombin III, and tissue factor pathway inhibitor [TFPI]) whose deficiency results in thrombosis, α1-antitrypsin whose deficiency results in emhysima, and LDL-receptor whose deficiency results in hypercholestrolemia.

Yet other examples of a nucleotide sequence of interest include those encoding enzymes involved in specific metabolic defects (*e.g.*, urea cycle enzymes, especially ornithine transcarbamylase, argininosuccinate synthase, and carbamyl phosphate synthase); receptors (*e.g.*, LDL receptor); toxins; thymidine kinase to ablate specific cells or tissues; ion channels (*e.g.*, chloride channel of cystic fibrosis); membrane transporters (*e.g.*, glucose transporter); and cytoskeletal proteins (*e.g.*, dystrophin). The nucleotide sequence of interest may be of synthetic, cDNA, or genomic origin, or a combination thereof. The nucleotide sequence of interest may be one which occurs in nature, a non-naturally occurring gene which nonetheless encodes a naturally occurring polypeptide, or a gene which encodes a recognizable mutant of such a polypeptide. It may also encode an mRNA which will be "antisense" to a DNA found or to an mRNA normally transcribed in the host cell, but which antisense RNA is not itself translatable into a protein. In one embodiment, the invention discloses transgenic mice which express in a liver-specific manner the exemplary hFIX coding sequence under the control of AE5' (Example 3).

The term "liver-specific manner" as used herein in reference to the expression of a nucleotide sequence of interest in a transgenic animal is a relative term which means that the quantity of mRNA and/or protein encoded in liver tissue by the nucleotide sequence of interest is greater than, preferably two times greater, more preferably five times greater, and most preferably ten times greater, than the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest in tissues other than liver tissue of the same transgenic animal as detected by Northern blot hybridization and/or by the activity of the encoded protein as described herein. Thus, the term "liver-specific" when made in reference to expression of a

nucleotide sequence of interest by a transgenic animal means that the transgenic animal expresses the nucleotide sequence of interest in an liver-specific manner.

A first step in the generation of the transgenic animals of the invention is the introduction of a construct containing nucleic acid sequences of interest under the expression regulatory control of sequences of the invention into target cells. Several methods are available for accomplishing this, including microinjection, retroviral infection, and implantation of embryonic stem cells. These methods are discussed as follows.

i. Microinjection Methods

Direct microinjection of expression vectors into pronuclei of fertilized eggs is the preferred, and most prevalent, technique for introducing heterologous nucleic acid sequences into the germ line. Technical aspects of the microinjection procedure and important parameters for optimizing integration of nucleic acid sequences have been previously described [Hogan *et al.*, (1986) Manipulation of the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Lab.].

Once the expression vector has been injected into the fertilized egg cell, the cell is implanted into the uterus of a pseudopregnant female and allowed to develop into an animal. Of the founder transgenic animals born, 70% carry the expression vector sequence in all of their cells, including the germ cells. The remaining 30% of the transgenic animals are chimeric in somatic and germ cells because integration of the expression vector sequence occurs after one or more rounds of replication. Heterozygous and homozygous animals can then be produced by interbreeding founder transgenics. This method has been successful in producing transgenic mice, sheep, pigs, rabbits and cattle [Hammer *et al.*, (1986) J. Animal Sci.:63:269; Hammer *et al.*, (1985) Nature 315:680-683].

ii. Retroviral Methods

Retroviral infection of pre-implantation embryos with genetically engineered retroviruses may also be used to introduce transgenes into an animal cell. For example, blastomeres have been used as targets for retroviral infection [Jaenisch, (1976) Proc. Natl. Acad. Sci USA 73:1260-1264]. Transfection is typically achieved using a replication-defective retrovirus carrying the transgene [Jahner *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82:6927-6931; Van der Putten *et al.*, (1985) Proc. Natl. Acad Sci USA 82:6148-6152].

Transfection is obtained, for example, by culturing eight-cell embryos, from which the zona pellucida has been removed with fibroblasts which produce the virus [Van der Putten (1985), *supra*; Stewart *et al.*, (1987) EMBO J. 6:383-388]. The transfected embryos are then transferred to foster mothers for continued development. Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele [Jahner *et al.*, (1982) Nature 298:623-628]. Yet another alternative method involves intrauterine retroviral infection of the midgestation embryos [Jahner *et al.* (1982), *supra*].

The advantages of retroviral infection methods include the ease of transfection and the insertion of a single copy of the transgene, which is flanked by the retroviral long terminal repeats (LTRs), into the chromosome. However, this method is not a preferred method because most of the founders will show mosaicism since infection occurs after cell division has begun. This necessitates outbreeding to establish homozygous and heterozygous lines suitable for analysis of gene expression. More importantly, the retroviral LTR sequences may interfere with the activity of the hINV upstream sequences in directing expression of the heterologous nucleic aid sequences.

iii. Embryonic Stem Cell Implantation

Another method of introducing transgenes into the germ line involves using embryonic stem (ES) cells as recipients of the expression vector. ES cells are pluripotent cells directly derived from the inner cell mass of blastocysts [Doetchman *et al.*, (1988) Dev. Biol. 127:224-227], from inner cell masses [Tokunaga *et al.*, (1989) Jpn. J. Anim. Reprod. 35:173-178], from disaggregated morulae [Eistetter, (1989) Dev. Gro. Differ. 31:275-282] or from primordial germ cells [Matsui *et al.*, (1992) Cell 70:841-847]. Expression vectors can be introduced into ES cells using any method which is suitable for gene transfer into cells, *e.g.*, by transfection, cell fusion, electroporation, microinjection, DNA viruses, and RNA viruses [Johnson *et al.*, (1989) Fetal Ther. 4 (Suppl. 1):28-39].

The advantages of using ES cells include their ability to form permanent cell lines *in vitro*, thus providing an unlimited source of genetic material. Additionally ES cells are the most pluripotent cultured animal cells known. For example, when ES cells are injected into an intact blastocyst cavity or under the zona pellucida, at the morula stage embryo, ES cells are capable of contributing to all somatic tissues including the germ line in the resulting chimeras.

Once the expression vector has been introduced into an ES cell, the modified ES cell is then introduced back into the embryonic environment for expression and subsequent transmission to progeny animals. The most commonly used method is the injection of several ES cells into the blastocoel cavity of intact blastocysts [Bradley *et al.*, (1984) Nature 309:225-256]. Alternatively, a clump of ES cells may be sandwiched between two eight-cell embryos [Bradley *et al.*, (1987) in "Teratocarcinomas and Embryonic Stem Cells: A Practical Approach," Ed. Robertson E.J. (IRL, Oxford, U.K.), pp. 113-151; Nagy *et al.*, (1990) Development 110:815-821]. Both methods result in germ line transmission at high frequency.

Target cells which contain the heterologous nucleic acid sequences are recovered, and the presence of the heterologous nucleic acid sequence in the target cells as well as in the animal is accomplished as described *supra*.

E. Gene Therapy

The regulatory nucleic acid sequences provided herein may be used for gene therapy applications in both non-human animals as well as in humans. For example, the regulatory nucleic acid sequences of the invention may be introduced into cells using an expression vector which encodes a polypeptide sequence of interest using a variety of means known in the art to be useful both for delivery *in vivo* and *ex vivo*, including (1) recombinant retroviral transduction, (2) recombinant adenoviral vectors, (3) targeted cationic liposomes, and (4) gene transfer using biolistics, as described in the following sections.

1. Recombinant Retroviral Transduction

Retroviral vectors encoding polypeptides of interest may be used for the expression of the polypeptides in any desired cell, such as primary tumor cells. The transfer of polypeptides of interest using retroviruses may be made more efficient by increasing the titer of the virus encoding the polypeptides of interest and increasing the transduction efficiency. To increase the virus titer, the retroviral construct may be designed to include a selectable marker (*e.g.*, *neo* gene), and cells harboring the retroviral construct are selected by growth in the presence of a suitable selective agent (e.g., G418) followed by expansion of clones producing the highest titers of virus. To improve the transduction efficiency, retrovirus are

used in combination with liposomes or poly-L-ornithine or polylysine to enhance virus uptake.

Another way to improve gene transfer efficiency using retroviruses is to increase the targeting efficiency. Many tumor cells including glioblastomas and melanomas express excess levels of the transferrin receptor. Transferrin has been used to increase the transduction efficiency of adenovirus in combination with polylysine. Several recent reports demonstrated that replacing the SU (surface) domain of the *env* gene of a retrovirus can increase receptor-mediated transduction efficiency. The human transferrin gene is 2097 bp long and its insertion into the SU domain of the *env* gene of MLV vector may not produce a stable Env product. However, since earlier studies have suggested that the modified Env fusion protein requires the native Env for stable assembly and efficient entry, co-transfection of the transferrin-*env* fusion gene with the native *env* gene may be used to produce retrovirus particles bearing a mixture of wild type and recombinant Env. The gene transfer efficiency of the new vector may be examined by transducing tumor cells expressing high levels of transferrin receptor.

2. Recombinant Adenoviral Vectors

Recombinant adenoviruses can accommodate relatively large segments of foreign DNA (~7 kb), and have the advantage of a broad host cell range and high titer virus production. Adenoviruses have been used *in vivo* in rats to efficiently deliver genes to the liver and the pancreatic islets [reviewed in Becker *et al.* (1994) In *Protein Expression in Animal Cells*, Roth *et al.* eds.] and to the central nervous system [Davidson *et al.* (1993) Nature Genet. 3:219]. Rat livers have also been efficiently transduced *ex vivo* and then re-implanted [Shaked *et al.* (1994) Transplantation 57:1508]. Thus, the present invention contemplates *ex vivo* transfection followed by transplantation of the transfected cells or organ.

The replication defective recombinant adenoviruses are preferably employed; these viruses contain a deletion of the key immediate early genes E1a and E1b. To generate and propagate recombinant viruses, a packaging cell line such as 293 cells which supply the E1a and E2a proteins *in trans* is employed. Recombinant adenoviruses are created by making use of intracellular recombination between a much larger plasmid encoding most of the viral genome and a small plasmid containing the nucleotide sequence of interest flanked by

regions of homology with the viral integration site. Standard methods may be used to construct the recombinant adenoviruses [Graham and Prevec (1991) Meth. Mol. Biol. 7:109-128; Becker *et al.* (1994) In *Protein Expression in Animal Cells*, Roth *et al.* eds.]. Briefly, each plasmid is co-transfected together with pJM17 (Microbix Systems, Toronto) into sub-confluent monolayers of 293 cells (ATCC CRL 1573) using calcium phosphate precipitation and a glycerol shock. Initial recombinant viral stocks are titered on monolayers of 293 cells, and isolated single plaques are obtained and tested for expression of the polypeptide of interest using ELISA. Viral stocks are amplified and titered on 293 cells, and stored in aliquots at -70°C; if necessary, stocks are concentrated by centrifugation on density gradients. To infect tumor cells with recombinant adenoviruses, freshly isolated tumor cells are mixed with adenoviral stocks in a minimal volume. Titers of stocks are typically 10^5-10^8/ml. Medium is replaced after several hours and the cells are followed for expression of the recombinant adenoviral-encoded polypeptide of interest (*e.g.*, reporter genes).

A potential drawback of using an adenoviral delivery system is that the transduced cells may retain or express small quantities of adenoviral antigens on their surface. "Second generation" adenoviral vectors which contain deletions in the E2a gene are available and are associated with less inflammation in the recipient and a longer period of expression of the gene of interest [Engelhardt *et al.* (1994) Proc. Natl. Acad. Sci. USA 91:6196]. If necessary, nucleic acid sequences encoding polypeptides of interest are inserted into second generation adenoviral vectors.

Recently, adenoassociated virus (AAV) vectors and chimeric lentivirus vectors have also been shown promise in the expression of polypeptide sequences of interest.

3. Targeted Cationic Liposomes

Cationic liposomes have proven to be a safe and effective means for inducing the transient expression of DNA in target cells [Ledley (1995) Human Gene Ther. 6:1129]. Clinical trials are underway using cationic liposomes to introduce the CFTR gene into the lungs of cystic fibrosis patients [Caplen *et al.* (1994) Gene Ther. 1:139 and Alton *et al.* (1993) Nature Genet. 5:135] or to introduce, by direct intra-tumor injection, the T cell costimulator B7-1 into malignant melanoma lesions in order to induce a cell-mediated immune response [Nabel *et al.* (1993) Proc. Natl. Acad. Sci. USA 90:11307].

Cationic liposomes (*e.g.*, DOTAP/DOPE) and ligand-targeted cationic liposomes may be employed for the delivery of polypeptides of interest to tumor cells. Recently, in addition to cationic liposomes, neutral liposomes have also been reproted to also be useful in targeing ligands to cells. Ligand-targeted liposomes are made by covalently attaching ligands or antibodies to the surface of the cationic liposome. For example, when glioblastoma cells are to be targeted, transferrin is used as the ligand as glioblastoma cells express high levels of the transferrin receptor on their surface. When melanoma cells are to be targeted, internalizing receptors, monoclonal antibodies directed against melanoma-specific surface antigens (*e.g.*, mAb HMSA5) may be employed as the ligand.

Plasmid DNA encoding polypeptides of interest is formed into a complex with preformed cationic liposomes using standard methodology or alternatively the DNA is encapsulated into the liposome interior. The DNA-containing liposomes are then used to transfer the DNA to tumor cells *in vivo* by direct intra-tumor injection or *in vitro* (using freshly explanted tumor cells) followed by return of the transduced cells to the recipient (*e.g.*, a human patient or non-human animal).

4. Gene Transfer Using Biolistics

Biolistics (microballistics) is a method of delivering DNA into cells by projection of DNA-coated particles into cells or tissues. DNA is coated onto the surface of gold or tungsten microparticles (~1-3 μm diameter) and these particles are accelerated to high velocity and are impacted onto the target cells. The particles burst through the cell membrane and lodge within the target cell. The cell membrane quickly reseals and the passenger DNA elutes off of the particle and is expressed. The biolistic method has been used to transfect mammalian cells [Sanford *et al.* (1993) Methods Enzymol. 217:483].

A hand-held biolistic apparatus (BioRad) is used to transfer DNA into tumor cells or isolated tumor fragments. This device uses compressed helium to drive a disc-shaped macroprojectile which carries on its surface microparticles (1-5 μm) of gold which have been coated with purified plasmid DNA (coprecipitated with spermine) (Williams *et al.*, *supra*). This apparatus has been used to successfully transfect primary tissues.

Plasmid DNA encoding the polypeptides of interest may be coated onto the surface of gold microparticles according to the manufacturer's instructions (BioRad) and the biolistic apparatus is used to transfer the DNA into freshly explanted tumor cells or directly into

exposed tumors (*e.g.*, metastatic nodules on the surface of the liver, melanoma lesions on the skin).

Regardless of the method of delivery of the expression vector into a cell, it is preferred, though not required, that the expression vector contain a selection marker (*e.g.*, *neo* gene) to facilitate selection of transfected cells. Transfected cells are selected by growth in the presence of G418 (*e.g.*, 200 μg/ml), followed by culture in growth medium containing reduced concentrations of G418 (*e.g.*, 100 μg/ml) and growth to confluence. Expression of the polypeptides of interest is evaluated using, for example, immunoblot analysis or flow cytometry using monoclonal antibodies which are specific for the polypeptides of interest. It is preferred, though not necessary, that expression of the polypeptides of interest in the transfected tumor cells is both constitutive and stable. Constitutive expression refers to expression in the absence of a triggering event or condition, and can be achieved by the selection of a promoter which drives expression of the nucleic acid sequence encoding the polypeptides of interest. Examples of promoters which drive constitutive expression of a structural nucleic acid sequence which is operably linked to the promoter include the SRα promoter, CMV promoter, and HIV promoter.

Regardless of the type of expression vector used for delivery of the nucleic acid sequences of interest into a cell, the expression vector may be introduced to the cell by direct injection into tumor and/or preneoplastic tissue, systemic (*e.g.*, intravenous) administration, aerosol administration (e.g., for delivery to the bronchial tree and other lung tissues), injection into breast ducts (*e.g.*, for delivery to breast tissue), and topical administration (*e.g.*, for delivery to cervical tissue).

F. Reducing Expression Of Factor IX In An Animal

The regulatory sequences of the invention may also be used to reduce expression of a polypeptide sequence of interest which is encoded by a nucleic acid sequence whose transcription is under the regulatory control of the regulatory sequences provided herein. For example, the regulatory sequences of the invention may be used to reduce the rate of age-related increase of FIX activity in an animal as a means of treating diseases (*e.g.*, thrombosis, cardiovascular disease, *etc.*) which are associated with age-related increases in FIX activity. Since the inventors have discovered that the exemplary nucleic acid sequences AE5' and AE3' regulate stable and increased expression levels, respectively, of hFIX, the

increase in the level of hFIX activity over time may be reduced by inhibiting the function of AE3' which regulates increased expression of hFIX. This approach has the advantage that expression of hFIX remains under the control of AE5' thus providing hFIX activities which are stable over time and which continue to play an important role in normal blood coagulation processes.

The function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of the protein which specifically binds to AE3'. The protein(s) which bind to AE3' may be identified by using the AE3' (or the minimum portion of AE3' which has age-related regulatory activity) to screen protein libraries for specific binding to AE3' or its portion. Once the protein which binds to AE3' is identified, the function of this protein may be inhibited using antibodies which are specific for this protein. Antibodies which are specific for the protein which binds to AE3' are expected to disrupt the interaction between AE3' and this protein.

Antibodies (polyclonal and monoclonal) which are specific for the protein that binds to AE3' or portions thereof may be generated using methods known in the art. The term "antibody" refers to immunoglobulin evoked in animals by an immunogen (antigen). It is desired that the antibody demonstrates specificity to epitopes contained in the immunogen. The term "polyclonal antibody" refers to immunoglobulin produced from more than a single clone of plasma cells; in contrast "monoclonal antibody" refers to immunoglobulin produced from a single clone of plasma cells. The terms "specific binding," "specifically binding" and grammatical equivalents thereof when used in reference to the interaction of an antibody and an immunogen mean that the interaction is dependent upon the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the immunogen; in other words the antibody is recognizing and binding to a specific immunogen structure rather than to immunogens in general. For example, if an antibody is specific for epitope "A", the presence of an immunogen containing epitope A (or free, unlabelled A) in a reaction containing labelled "A" and the antibody will reduce the amount of labelled A bound to the antibody.

Polyclonal and monoclonal antibodies which are specific to a desirable polypeptide, given the teachings herein, may readily be prepared by one of skill in the art. For example, monoclonal antibodies may be generated by immunizing an animal (*e.g.*, mouse, rabbit, *etc.*) with a desired antigen and the spleen cells from the immunized animal are immortalized,

commonly by fusion with a myeloma cell. Immunization with antigen may be accomplished in the presence or absence of an adjuvant, *e.g.*, Freund's adjuvant. Typically, for a mouse, 10 μg antigen in 50-200 μl adjuvant or aqueous solution is administered per mouse by subcutaneous, intraperitoneal or intra-muscular routes. Booster immunization may be given at intervals, *e.g.*, 2-8 weeks. The final boost is given approximately 2-4 days prior to fusion and is generally given in aqueous form rather than in adjuvant.

Spleen cells from the immunized animals may be prepared by teasing the spleen through a sterile sieve into culture medium at room temperature, or by gently releasing the spleen cells into medium by pressure between the frosted ends of two sterile glass microscope slides. The cells are harvested by centrifugation (400 x g for 5 min.), washed and counted. Spleen cells are fused with myeloma cells to generate hybridoma cell lines. Several mouse myeloma cell lines which have been selected for sensitivity to hypoxanthine-aminopterin-thymidine (HAT) are commercially available and may be grown in, for example, Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) containing 10-15% fetal calf serum. Fusion of myeloma cells and spleen cells may be accomplished using polyethylene glycol (PEG) or by electrofusion using protocols which are routine in the art. Fused cells are distributed into 96-well plates followed by selection of fused cells by culture for 1-2 weeks in 0.1 ml DMEM containing 10-15% fetal calf serum and HAT. The supernatants are screened for antibody production using methods well known in the art. Hybridoma clones from wells containing cells which produce antibody are obtained, *e.g.*, by limiting dilution. Cloned hybridoma cells (4-5 x 10^6) are implanted intraperitoneally in recipient mice, preferably of a BALB/c genetic background. Sera and ascites fluids are collected from mice after 10-14 days.

The invention also contemplates humanized antibodies which may be generated using methods known in the art, such as those described in U.S. Patent Numbers 5,545,806; 5,569,825 and 5,625,126, the entire contents os which are incorporated by reference. Such methods include, for example, generation of transgenic non-human animals which contain human immunoglobulin chain genes and which are capable of expressing these genes to produce a repertoire of antibodies of various isotypes encoded by the human immunoglobulin genes.

Alternatively, the function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of AE3' using antisense sequences which are directed to AE3'. The term "antisense" as used herein refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a strand of a DNA duplex. AE3' antisense sequences may be used to turn off genes under the expression regulation of AE3' by transfecting a cell or tissue with expression vectors which express high levels of a desired AE3' antisense oligomer (*e.g.*, 15-20 nucleotides) or larger fragment. Such constructs can flood cells with antisense sequences which inhibit expression of FIX. Antisense sequences can be designed from various locations along the AE3' sequence. Animals (*e.g.*, mice) treated with vectors expressing AE3' antisense sequences are monitored for changes in the age-related symptoms associated with FIX expression. The alleviation or treatment of one or more of these symptoms in animal by an antisense sequence suggests that the antisense sequence may be useful in the treatment and/or prevention of age-related FIX expression in humans.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof. Unless otherwise mentioned, all reference to nucleotide numbers with respect to the factor IX nucleotide sequence, refers to the nucleotide numbers of the hFIX gene sequence shown in Figure 8.

EXAMPLE 1

Construction Of A Series of Twelve Exemplary Human Factor IX (hFIX) Minigene Expression Vectors

To explore the molecular mechanisms underlying age-related regulation of Factor IX, a series of twelve hFIX minigene expression vectors were constructed. These vectors were first analyzed *in vitro* in HepG2 cells, a human hepatoma cell line (see Example 2, *infra*). Transgenic mice harboring the hFIX minigene vectors were generated and longitudinal analyses of hFIX expression for the entire life spans of founders and successive generations of transgenic mice were carried out (See Example 3, *infra*).

The twelve exemplary minigenes contained sequences derived entirely from the hFIX gene sequence, including (a) promoter sequences of various lengths spanning up to nucleotide (nt) -2231 in the 5′ flanking region, (b) the coding region containing a first intron in which the first intron's middle portion is truncated. *i.e.*, nt +1098 through nt +5882 of Figure 8, and (c) either the complete 3' UTR sequence or the 3' UTR sequence in which the middle portion was deleted. Figure 1 shows the structure of eleven out of the twelve human FIX minigene expression constructs. The name of each construct is shown at left. The structure is depicted with the promoter-containing regions (solid thick line on left) with the 5′ terminal nucleotide number. Transcribed hFIX regions (open rectangles connected with thin lines representing the shortened first intron) are followed by 3′ flanking sequence regions (solid thick line at right). Arrow: transcription start site; asterisk: translation stop codon; pA: polyadenylation; sl: potential stem-loop forming dinucleotide repeats.

Construction of hFIX minigene expression vectors was carried out using -416FIXm1 as the starting construct (Kurachi *et al.* (1995) J. Biol. Chem. 270:5276-5281). The nucleotide (nt) numbering system used in this study was based on the complete hFIX gene sequence previously reported (Yoshitake *et al.* (1985) Biochem. 24:3736-3750). Minigene -416FIXm1/1.4 was constructed from -416FIXm1 by inserting the middle portion of the 3′ UTR (1.2 kb in size) which was generated by PCR using the following primer set with BamH I linkers: 5′ primer, TAACAGGATCCGGCCTCTCACTAACTAATCAC (nt +31418 through +31438) (SEQ ID NO:14) and 3′ primer, CAACTGGATCCAAGATTCAAGATAGAAGGAT (nt +32690 through +32671) (SEQ ID NO:15), and human genomic DNA as an amplification template. The PCR product was digested with BamH I, and the generated fragment was inserted into the 3′ UTR BamH I site of -416FIXm1, thus producing -416FIXm1/1.4 which contained the entire 3′ UTR. -416FIX m1/0.7 was constructed by inserting the PCR-amplified 700 bp fragment with BamH I linker, containing the 3′ contiguous sequence to nt +32117. The primers used were, 5′ primer: same as that for -416FIXm1/1.4, 3′primer: GGACAGGATCCCC CAAACTTTTCAGGCAC (nt +32117 through +32097) (SEQ ID NO:16). Minigenes -590FIXm1, -679FIXm1, -770FIXm1, -802FIXm1 and -2231FIXm1 were produced by replacing the 5′ end 433 bp sequence of -416FIXm1 released by Sph I/Nhe I digestion with 607, 696, 787, 819 and 2248 bp fragments containing the 5′ end hFIX region extended up to nt -590, -679, -770, -802 and -2231, respectively. These latter fragments were generated by

Sph I/Nhe I digestion of the PCR product obtained with 5′ primers: CAAGCATGCATCTAGTGTTAGTGGAAGAG (nt -590 through -571) (SEQ ID NO:17), CAAGCATGCAAATATTAACTCAAAATGGA (nt -679 through -660) (SEQ ID NO:18), CAAGCATGCTGTTGTTTTTTGTTTTAAC (nt -770 through -752) (SEQ ID NO:19), CAAGCATGCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:20), CAAGCATGCGATCCCTTCCTTATACCT (nt -2231 through -2214) (SEQ ID NO:21) with Sph I linker and the common 3′ primer TAAGCTTAACCTTTGCTAGCAGATTGT (nt +30 through +10) (SEQ ID NO:22) and human genomic DNA as the amplification template. Minigene -802FIXm1/0.7 (whose structure is not shown in Figure 1) contains the 3' UTR region through nt 32,140, which is then connected to nt 32,690 through its downstream poly (A) signal sequence that is common to each of the other eleven constructs.

-416FIXm1/AE5′ depicts a construct with the AE5′ region moved to the 3′-end position and shown as an open box at right. -416FIX m1/AE5′ was constructed by inserting the Kpn I fragment generated by PCR (nt -802 through nt -417) into the -416FIXm1 vector (the Kpn I site is outside of the FIX gene, Figure1). The 5′ and 3′ primers used for PCR were CTTGGTACCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:23) and CTTGGTACCATATGAATCCTTTCATAGAT, (nt -417 through -436) (SEQ ID NO:24) respectively. All constructs were sequenced through PCR amplified regions as well as fragment ligation sites to confirm the correct sequences and orientations.

EXAMPLE 2

Transient Expression of Eleven hFIX Minigene Expression Vectors *In Vitro* In Human Hepatoma HepG2 Cell Line

Transient *in vitro* expression activities of hFIX minigene constructs were assayed using HepG2 cells and hFIX specific enzyme linked immunosorbent assay (ELISA) as previously described (Kurachi *et al.* (1995) J. Biol. Chem. 270:5276-5281) with some modifications. Cell transfection was carried out by the calcium phosphate-DNA conjugate method or later, using FuGene 6 (Boehringer Manheim). The latter, improved transfection method consistently increased transfection efficiency to >20% (Kurachi *et al.* (1998) Biochemica 3:43-44), and all earlier assays were reexamined using FuGene 6. Four to five independent assays of factor IX activity were carried out and the averages were shown with

standard errors. With FuGene 6 transfection, the control minigene -416FIXm1 typically produced hFIX at a level of ~50 ng/10^6 cells/48 hr.

Figure 1 shows the relative *in vitro* transient expression activities of the human FIX minigene expression constructs (transient expression activity of minigene -802FIXm1/0.7 which is not shown in Figure 1 was 81.5% of the acitity of minigene -416FIXm1). Transient expression activities relative to the activity of -416FIXm1 (~50 ng/10^6 cells/48 hour, and defined as 100% activity) are shown on the right side with standard deviations (from 4-5 independent assays). Activities were normalized to the size of minigenes used.

The relative transient expression activities shown in Figure 1 show that all constructs showed comparable high transient hFIX expression in HepG2 cells (~50 ng/10^6 cells/48 hours). However, all the constructs containing the complete 3′ UTR, including a 102 base pair (bp) stretch of inverted AT, GT and GC dinucleotide repeats [Yoshitake *et al.* (1985) Biochem. 24:3736-3750], reproducibly showed expression activity levels which were 25-30% lower than corresponding minigenes without the repeat sequences. Dinucleotide repeats similar to those seen in the hFIX 3′ UTR, which can form stable stem-loop (sl) structures in mRNA, have been implicated in controlling mRNA stability in mammals as well as yeast and plants, thus providing an important layer of protein biosynthesis regulation [Ross (1995) Microbiol. Rev. 59:423-450]. Together, these results suggest a similar negative regulatory activity for this structure of the hFIX gene in the HepG2 assay system on expression of the hFIX gene. As described below (*e.g.*, Example 3), however, the 3′ UTR structure of the hFIX gene containing the dinucleotide repeat region showed unexpected functions *in vivo* which are critical for advancing age-related regulation of the hFIX gene.

Another important and surprising finding with the HepG2 cell assay system is that expression by these hFIX minigenes (which contained sequences which are positioned upstream and downstream of the hFIX gene, and which are derived from the homologous hFIX gene instead of from heterologous reporter genes) does not show any down-regulation in the presence of the 5′ upstream region (nt -802 up through nt -1900) [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068] (Figure 1). In contrast, when a CAT reporter gene was used, negative regulatory elements were identified in this region [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068].

EXAMPLE 3

Generation And Analysis Of Transgenic Mice Harboring hFIX Minigene Expression Vectors

Transgenic animals were constructed using the expression plasmids described above in Example 2 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition). All animal experiments were carried out in accordance with the institutional guidelines of the University of Michigan (OPRR No. A3114-01).

Briefly, Factor IX minigene expression plasmids were double-digested with Sph I/Kpn I and the factor IX minigene-containing fragments released were isolated by 0.8% agarose gel electrophoresis, followed by purification with SpinBind DNA extraction units (FMC). Fertilized eggs of C57B/6 X SJL mice were microinjected with the DNA (1-2 ng/egg), and implanted into foster mother animals (CD-1).

A. Multiplex PCR Analysis

Offspring produced were screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples. Two pairs of primers were used, one specific to the hFIX transgenes and the other specific to mouse β-globin gene (endogenous control); 5' primer: CTGTGGGAACACACAGATTTTGG (nt +6172 through +6195) (SEQ ID NO:25) and 3' primer: GGAATAATTCGAATCACAT (nt +30885 through +30867) (SEQ ID NO:26), and 5' primer: CCAATCTGCTCACACAGGAT (nt +2590 through +2609) (SEQ ID NO:27) and 3' primer: CCTTGAGGCTGTCCAAGTGA (nt +3083 through +3064) (SEQ ID NO:28), respectively. These primers were designed to amplify a unique 966 bp fragment from the hFIX transgenes and a 494 bp fragment from the mouse β-globin gene, respectively. PCR was initiated with 3 min incubation at 94° C, followed by 25 cycles of 94° C for 30 sec, 65° C annealing for 1 min and 72° C extension for 2 min.

Founders were back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals were generated by crossing among heterozygous F1 littermates and the following generations were similarly generated. Zygosity status of animals was determined by quantitative multiplex PCR analysis as described above.

Minimally, three founder lines for each minigene construct were subjected to longitudinal analysis for their entire life spans up to two years.

Figure 3B shows the results of quantitative multiplex PCR analysis to determine the relative transgene levels in tail and liver tissues. Genomic DNA was extracted from snipped tail tissue of a transgenic -416FIXm1 animal (PA112) at 3 weeks and at 19 months of age. Liver DNA was extracted from the same animal (PA112) sacrificed at 19 months of age and a -416FIXm1 animal (PA412) sacrificed at 1 month of age. Positions of hFIX specific fragment (966 bp) and mouse β-globin specific fragment (494 bp, internal copy number control) are shown on the right. Lane 1: kb size ladder ; lane 2: fragment size control amplified from -416FIXm1 plasmid; lane 3: non-transgenic mouse tail DNA as template; lane 4: tail DNA of PA112 at 3 weeks of age; lane 5: tail DNA of PA112 at 19 months of age; lane 6: liver DNA of PA412 at 1 month of age; lane 7: liver DNA of PA112 at 19 month of age. The relative transgene copy numbers for the 1 month-old versus the 19 month-old animals, normalized to the endogenous mouse β-globin gene, were 1.0-1.1 for both tail as well as liver genomic DNA preparations, showing no sign of loss of the hFIX transgene in the genome with age (Multi Analyst program from BioRad used for quantitation and calculation of ratios).

B. Immunoassay of hFIX Levels In Transgenic Mice

Circulatory hFIX levels were monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying various hFIX minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection (aliquot of ~100 μl) via tail-tip snipping, and the obtained serum was routinely used to quantify hFIX levels in the circulation using duplicated hFIX-specific ELISA for each age point. Pooled human plasma (George King Bio-Medical) was used to prepare a hFIX standard curve for each assay. In order to minimize experimental fluctuations from assay to assay in the longitudinal analysis, overlapped serum samples from the previous assay group were included in each assay. To ensure reproducibility, three to six independent founder lines were generated for each minigene construct, and animals from at least three representative lines were subjected to longitudinal analyses for their entire life spans. The duplicated ELISA values varied less than 11% from the averages. The results are shown in Figures 2 and 4.

In all panels in Figures 2 and 4, labeling of animals is based on the tag numbers plus additional information. The first letters of the label F or P represent founder or progeny, respectively. Information on progeny generation (F1 or F2) and sex are in parenthesis (m: male; f: female), followed by status (+: alive in good health; d: died; s: sacrificed for various examinations; mo: age of death or sacrifice). To avoid overcrowding of the panels, the results from representative animals are shown for each minigene construct. Importantly, age-regulation patterns were remarkably similar among all animals for each specific construct and different founder lines. Panels A-E of Figure 2 show representative founder line animals with -416FIXm1 (A); -416FIXm1/1.4 (B), -590 FIXm1 (C), -679FIXm1 (D) and -770FIXm1 (E). Panels A-D of Figure 4 show representative founder line animals with -802FIXm1, -802FIXm1/1.4, -2231 FIXm1 and -2231FIXm1/1.4, respectively. Panel E shows representative founder line animals with -416FIXm1/AE5'.

Figure 2 shows that at one month of age, the mice carrying the -416FIXm1 minigene produced hFIX at varying levels, from as high as that of natural hFIX gene expression (~4 μg/ml) to much lower levels (~50 ng/ml) (Figure 2A). Such variations are primarily due to the transgene positional effects in the genome. Circulatory hFIX levels of animals from the representative founder lines carrying the minigene, however, declined drastically through puberty and during the subsequent two to three month period to much lower levels, which then remained stable for the remaining life span. This rapid age-dependent characteristic decline in the circulatory hFIX level was observed in all animals analyzed (n=69), regardless of founder line, differences in initial hFIX level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, or zygosity status (homozygous / heterozygous) of the transgenes.

C. Northern Blot Analysis of hFIX mRNA in Transgenic Mice

Northern blot analyses of the liver RNA samples from animals (15 μg per lane) were carried out as previously described [Kurachi *et al.* (1995) *supra*] using the ^{32}P-labeled Ssp I/BamH I fragment (the 3' half of the hFIX coding region of the cDNA) as a probe, and employing stringent washing conditions. Under these conditions, the probe preferentially hybridized strongly with hFIX minigene mRNA bands (~1.7 kb) with little cross-hybridization with the mouse FIX mRNA bands (3.2 kb and 2.2 kb) [Yao *et al.* (1994) Gene Therapy 1:99-107]. To confirm the presence of equivalent amounts of RNA in each lane, the filters

previously hybridized with hFIX probe were stripped of probe and re-probed with the RNR18 cDNA (ribosomal RNA 18S). After completion of longitudinal analyses of animals from key founder lines for their entire life spans, the representative lines were subjected to embryo-freezing for preservation.

The results of Northern analysis of human FIX mRNA and transgene DNA levels in the livers of animals carrying -416FIXm1 are shown in Figure 3A. hFIX mRNA levels in the liver of young (PA412: F1/f, 1 month of age) and old (PA112; F1/f, 19 months of age) transgenic animals were analyzed by Northern blot analysis of total liver RNA. PA412 and PA112 animals were from the same litter produced by the founder FA661, and expressed 1252 and 1675 ng/ml circulatory hFIX at one month of age, respectively. PA112 was expressing 63.8 ng/ml serum hFIX at the time of sacrifice. Lane 1: non-transgenic mouse liver RNA; lane 2: transgenic PA412 liver RNA; lane 3: transgenic PA112 liver RNA. FIX and 18S on the left or right sides indicate the band position of hFIXm1 mRNA (~1.7 kb) and RNR18 (1.9 kb, ribosomal RNA), respectively.

Figure 3A shows that the decline in blood hFIX level observed in Figure 2 was correlated with a similar decline in the steady-state liver hFIX mRNA, which was not due to a loss of the hFIX transgene with age (Figure 3B). This was further supported by the fact that when 4-5 month old mice with much decreased hFIX levels had progeny, their pups depicted pre-pubertal high hFIX expression levels equivalent to those of their parents at the same time point (one-month of age).

Minigene vector -416FIXm1/1.4 is identical to -416FIXm1 except that -416FIXm1/1.4 contains the complete 3'UTR, including the dinucleotide repeat structure (102 bp in length) in its middle region [Yoshitake *et al.* (1985) Biochem. 24:3736-3750] (Figure 1). Transgenic mice with -416FIXm1/1.4 (n=48) (Figure 2B) showed pre-pubertal high and subsequent age-dependent decline in hFIX levels similar to those of -416FIXm1 (Figure 2A), although the decline was less steep and expression levels were stabilized at significantly higher levels than those observed for -416FIXm1 (Figure 2B).

These results indicate that, while the 102-bp sequence containing the dinucleotide repeat structure of hFIX 3' UTR reduces the age-related decline in expression of hFIX, the presence of the complete 3' UTR containing the extensive dinucleotide repeat structure nonetheless does not completely rescue hFIX expression from the age-decline observed in all

of these animals, regardless of founder line, initial pre-pubertal hFIX level, generation, sex, or zygosity status of the transgenes.

All animals carrying minigenes -590FIXm1and -679FIXm1 (a total of 25 and 26 animals subjected to longitudinal analysis, respectively) also showed an age-associated rapid decline in hFIX expression similar to that seen in animals carrying -416FIXm1 (Figure 2, C and D). Furthermore, hFIX expression levels in three independent founder animals generated to date carrying -770 FIXm1 also rapidly decreased over the puberty period in a similar pattern as the above minigenes (Figure 2E). These observations indicated that minigenes with the promoter region up to nt -770 contain the basic structural elements necessary for hFIX expression, but lack a structural element(s) which functions in age-associated stability of hFIX gene expression.

In contrast, striking and unexpected differences in hFIX expression patterns were observed with animals carrying the minigene -802FIXm1 (Figure 4A) as comp[ared to those carrying the minigene -416FIXm1 (Figure 2A). -802FIXm1 is composed of a vector frame identical to -416FIXm1, except that the 5′ end flanking sequence included was extended to nt -802 (Figure 1).

All animals with -802FIXm1, -2231FIXm1 and -416FIXm1/AE5' (panels A, C, E) exhibited stable expression throughout their life spans. Animals with -802FIXm1/1.4 and -2231FIXm1/1.4 (Figure 4 B, D) exhibited age-associated increases in hFIX expression levels. All animals maintained or increased stable circulatory hFIX levels regardless of founder line, initial expression levels at one month of age, sex, generation or zygosity status. Mice which died at much younger ages than their normal life expectancies are marked with d. The above results show that all animals from three independent founder lines obtained with -802FIXm1 (Figure 4A) showed characteristic differences in hFIX expression pattern from animals with -416FIXm1 (Figure 2A) and -416FIXm1/1.4 (Figure 2B).

The -802FIXm1 transgenic animals (n = 62) subjected to longitudinal analysis invariably showed age-stable plasma hFIX levels for their entire life spans, mostly up to 20-24 months of age. Age-stable circulatory hFIX levels were correlated with age-stable mRNA levels (Figure 5). These observations with -802FIXm1 were further supported by age-stable hFIX expression by mice carrying -2231FIXm1 (Figure 4C). Together, these results suggest that the structural element which is responsible for age-stable expression of the hFIX gene resides in the small region spanning nt -770 through -802. We designated

this small region "age-regulatory element in the 5′ end" (AE5′). This region contains a transcription factor PEA-3 nucleotide sequence (GAGGAAG: nt -784 through -790), which completely matches the consensus motif (C/G)AGGA(A/T)G [Martin *et al.* (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin *et al.* (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre *et al.*(1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. The function of AE5′ nucleotide sequence is position-independent as shown by age-stable hFIX expression by animals containing -416FIXm1/AE5′, in which AE5′ was moved to the 3′ end outside of the hFIX minigene (Figure 4E).

Since transgenes of -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 differ from the minigenes -802FIXm1 and -2231FIXm1 only by their promoters, the hFIX mRNA produced from all of these minigenes (an intron spliced form of FIXm1 RNA) was expected to produce identical hFIX protein. Thus, it was hypothesized that the age-dependent decline in the circulatory hFIX level observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1, but not with -802FIXm1 and -2231 FIXm1, must be due to an age-dependent decline in the transcriptional activity of the transgenes. This agrees with the facts that no significant changes with age in hFIX mRNA levels in the liver were observed for animals carrying -802FIXm1 (Figure 5, lanes 2 and 3), while advancing age-dependent declines in steady-state mRNA level were observed for -416FIXm1 (Figure 3A, lanes 2 and 3).

To further determine whether the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes, the effects of age on hFIX clearance from the circulation were tested as follows. Aliquots of plasma-derived hFIX preparation (5 μg/0.1 ml of PBS) were injected via tail vein into normal animals at 2, 9-10 and 19-23 months of age (n=3 per age group), which have the same genetic background as the transgenic mice (C57B/6 X SJL). The hFIX level in circulation was monitored by ELISA of collected blood samples (~50 μl aliquot) at 10 min, 2, 6, 12, 18, 24, 30, 36 and 48 hrs after protein injection. As expected, all animals of different age groups showed a typical bi-phasic clearance kinetics (two compartment distribution and clearance) with an initial rapid clearance phase (α-phase), followed by a slower clearance phase (β-phase). The results are shown in Table 1.

Table 1

Clearance Time of Human Factor IX in Mice

Age (months)	Clearance Time ($T_{1/2}$ of Human Factor IX)
2	16.8 ± 0.21
9-10	17.4 ± 0.55
19-23	16.9 ± 0.35

As shown in Table 1, very similar half clearance times were observed for all age groups tested. This agreed with our previous results (17.8 hours) for hFIX clearance in a different strain, BALB/c mice (2 months of age) (Yao *et al.* (1994) *supra*].

Furthermore, the results in Table 1 demonstrate that the hFIX turnover time from the circulation does not change significantly *in vivo* with increasing age, from youth (2 months), to middle age (9-10 months) to old age (19-23 months). These results further confirm that the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes.

It is important to note that in the *in vitro* HepG2 cell assay system, the presence or absence of AE5′ in the minigenes did not make any significant difference in hFIX expression from the hFIX minigenes (Figure 1, and Example 2, *supra*). In contrast, as mentioned above in this Example, the presence or absence *in vivo* of AE5′ makes a dramatic age-dependent difference in hFIX gene expression. This further demonstrates that *in vivo* longitudinal analysis is important for studying age-regulation of a gene.

Unlike the animals with -802FIXm1, mice with -802FIXm1/1.4 (which contains the complete 3′UTR) showed an advancing age-associated increase in the hFIX level in the circulation (n=48) (Figure 4 A and B). Thus, to determine whether this unexpected age-dependent increase in the circulatory hFIX level was directly correlated with an increased level of liver hFIX mRNA, Northern blot analyses of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 were conducted. The hFIX mRNA levels in the liver of 1-month (young) or 15-month (aged) mice carrying -802FIXm1 (mouse P327 or P552, respectively) and -802FIXm1/1.4 (mouse P32 and P697, respectively) are shown in Figure 5. These

animals were from the same litter produced by the founder F17549 for -802FIX m1 and F229 for -802FIXm1/1.4 (Figure 4 A and B). At the time of sacrifice, P552 and P697 were expressing 2200 and 1658 ng/ml of hFIX, respectively. The total liver RNA (15 μg from each animal was used for the Northern blot analysis performed as described in Figure 3A. Upper panel: probed with the Ssp I/BamH I fragment of hFIX cDNA; lower panel: rehybridized with RNR18 (ribosomal RNA) probe. Lane 1: non-transgenic mouse liver; lane 2: transgenic P327 liver RNA; lane 3: transgenic P552 liver RNA; lane 4: transgenic P32 liver RNA; lane 5: transgenic P697 liver RNA. PhosphorImager (Molecular Dynamics) was used for quantitation of mRNA levels (counts) and ratios of young versus old were calculated. Young and old animals carrying -802FIXm1 showed no significant differences in the mRNA level (the ratio of old over young: 0.92). In contrast, -802FIXm1/1.4 animals showed a substantial elevation in the mRNA level with older age (the ratio of old over young: 1.54).

These results (Figure 5, lanes 4 and 5) indicated the presence of another important age-regulatory nucleotide sequence, designated AE3′, which is located approximately in the middle of the 3′ UTR where an extensive stretch of dinucleotide repeating structures were contained. In the presence of AE5′, AE3′ clearly confers a crucial age-associated increase in hFIX expression. This conclusion was further supported by results obtained with -2231FIXm1 /1.4 (n=42) (Figure 4D). The unique concerted function conferred by the combination of AE5′ and AE3′ was again independent of founder line, initial expression levels at one month of age, sex, generation, or zygosity status of animals.

Interestingly, animals with sustained high hFIX levels in the circulation (approximately 1,500 ng/ml or higher) tended to die at a much earlier age than the expected life span (~2 years) (Figure 4 A, B, D). This happened to both males and females, but appears to be more frequent in males. Without limiting the invention to any particular mechanism, it is believed that since these transgenic mice have hFIX in addition to their own mFIX, they may be at an increased risk of lethal thrombosis compared to wild type mice which do not express the transgenes.

The above-described characterization of transgenic mice harboring hFIX transgenes demonstrates that (a) while the presence of AE5' *in vitro* in HepG2 cells did not affect hFIX gene expression, the presence of AE5' *in vivo* resulted in a dramatic age-dependent increased stability in hFIX gene expression, (b) the age-dependent decline in the circulatory hFIX level

observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 is directly correlated with the decrease in the steady-state mRNA level, which the inventors believe to be due to an age-dependent decline in the transcriptional activity of the transgenes, and (c) animals carrying -802FIXm1/1.4 shwoed a substantial elevation in the liver mRNA levels of hFIX with older age.

EXAMPLE 4

Footprint And Gel Electrophoretic Mobility Shift Analysis Of The Region From Nucleotides -665 To -805 of Human Factor IX

In order to make a preliminary determination of the region within AE5' which is involved in the function of AE5', footprint analysis and gel electrophoretic mobility shift assays were performed as follows.

A. Footprint Analysis

For footprint analysis of the region spanning from nt -665 through nt -805, the fragments used were amplified by PCR with the ^{32}P-labeled 5′ primer ATGGTTAACTGACTTACGAA (nt -833 through -814) (SEQ ID NO:29) and 3′ unlabeled primer GCTCCATTTTGAGTTAATATTTGTGT (nt -657 through -682) (SEQ ID NO:30). The nuclear extracts (NEs) from HepG2 human hepatoma cells and livers of young (1 month of age) and old (19 months of age) mice were prepared as previously reported [Kurachi et al. (1986) Biochemistry 33:1580-1591]. Various amounts of NEs (0, 100 and 150 μg) were incubated with the labeled fragments (30,000 CPM) for 1 hour on ice and subjected to DNase 1 digestion (0.5 unit) for 2 min at room temperature. The samples tested included those without NEs, with 100 μg and 150 μg of HepG2 cell NEs, with 100 μg and 150 μg NEs from old mice, and with 100 μg and 150 μg NEs from young mice. Major and minor footprints and apparent DNase hypersensitivity sites were observed.

Footprint analysis of the region nt -665 through -805 with aged mouse liver nuclear extracts showed a major footprint (nt -784 through -802), a minor foot print (nt -721 through -728) and an interesting DNase hypersensitive region (nt -670 through -714). With nuclear extracts from one month-old animals or HepG2 cells, no such clear footprints were observed.

B. Gel Electrophoretic Mobility Shift Assay

Gel electrophoretic mobility shift assay using mouse liver nuclear extracts from three different age groups was used. Nuclear extracts were prepared from 1, 5 or 19 month-old mouse livers (as described *supra*). Double stranded oligonucleotides containing a PEA-3 nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (TTCAGTCGAGGAAGGATAGGGT) (SEQ ID NO:31) were ^{32}P-labeled at the 5′ end to a specific activity of 1.9×10^9 cpm. Aliquots of the radio-labeled oligonucleotide (20,000 cpm) were incubated with 10 μg of NEs in the presence of 1 μg of poly dI-dC in DNA binding buffer for 20 min at room temperature and subjected to polyacrylamide gel electrophoresis (Kurachi et al. (1986) *supra*). In Figure 6A, Lane 1: without NEs; lane 2: with NEs of 1 month-old mice; lane 3: with NEs of 5 month-old mice; land 4: with NEs of 19 month-old mice; lane 5: with mouse brain NEs (positive control for PEA-3, showing a slightly higher size of shifted band).

Figure 6B shows the results of the competition assay for ^{32}P-labeled double stranded oligonucleotides containing the PEA-3 nucleotide sequence. A 100-fold excess unlabeled oligonucleotide described in the preceding paragraph or mutant oligonucleotide [TTCAGTCGGTTGGTGATAGGGT (SEQ ID NO:32) with mutated sequences underlined] was incubated with 10 μg of 19 month-old mouse liver NEs for 5 min followed by addition of ^{32}P-labeled oligonucleotides as described *supra*. Lane 1: without NEs; lane 2: with NEs; lane 3: with NEs and wildtype competitor; lane 4: with NEs and mutant competitor.

In agreement with the above results of footprinting, gel electrophoretic mobility-shift (bandshift) assays showed an increase in protein binding with the nuclear extracts of aged mice (19 months of age) (Figure 6A). Bandshifts were competitively reduced with excess amounts of oligonucleotides harboring the PEA-3 motif, but not with oligonucleotides harboring a mutant PEA-3 motif sequence (Figure 6B), thus confirming the presence of the PEA-3 motif in AE5'. This is the first time that the PEA-3 protein, which is a member of the Ets family of transcription factors and which has been shown to bind to nucleotide sequences [SEQ ID NO:40; SEQ ID NO:48; and SEQ ID NO:84] that are homologous to the PEA-3 nucleotide sequence within the AE5' region [Karim *et al.* (1990) Genes & Develop. 4:1451-1453; Nelsen *et al.* (1993) Science 261:82-86; Fisher *et al.* (1991) Oncogene 6:2249-2254], has been implicated in such a unique role in age-stable regulation of a gene.

Without limiting the invention to any particular mechanism, the PEA-3 nucleotide sequence in the hFIX gene appears to have been generated through evolutionary drift of a L1 sequence originally recruited presumably via its retrotransposition into the 5′ specific location. Modern retrotransposable L1 [Kazazian *et al.* (1988) Nature 332:164-166; Dombroski *et al.* (1993) Proc. Natl. Acad. Sci. USA 90:6513-6517; Minakami *et al.* (1992) Nucl. Acids Res. 20:3139-3145; Dombroskiet *et al.* (1994) Mol. Cell. Biol. 14:4485-4492] does not have the corresponding PEA-3 nucleotide sequence. The PEA-3 nucleotide sequence of AE5′ nucleotide sequence resides within the L1-derived sequence retaining a 63-70% similarity with the ORF2 corresponding region of the modern retrotransposable L1 in the 5′ to 3′ orientation. Interestingly, the murine FIX gene also has the L1-derived nucleotide sequence in its 5′ end region in an almost identical position as in the hFIX gene, and has multiple PEA-3 consensus nucleotide elements [Kawarura *et al.*in *Organization of L1 Sequence in the 5' Flanking Region of Factor IX Gene* [in preparation]. Age-regulation of the murine FIX gene is indeed very similar to that of the hFIX gene [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi *et al.* (1996) Thromb. Haemost. 76:965-969], thus providing further insights into the evolutionary origin of the molecular mechanisms underlying age-associated regulation of the FIX gene.

EXAMPLE 5

Liver-Specific Expression Of The Exemplary hFIX Gene Under Control Of The hFIX Promoter

Expression of the natural FIX gene is virtually restricted to the liver [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068]. In order to determine whether any of the upstream and/or downstream sequences in the hFIX minigenes directed liver-specific expression of the hFIX transgene, Northern blot analysis was carried out as described *supra* (Example 3) in transgenic mice carrying -416FIXm1 and -802 FIXm1 expression vectors. Animals expressing hFIX at high level (PA412 and P580 carrying -416FIXm1 and -802FIXm1, respectively) were sacrificed at one month of age and total RNA was extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The results in transgenic mice carrying -416FIXm1 and -802 FIXm1 are shown in Figure 7A and B, respectively.

In Figure 7, the positions of hFIX mRNA, RNR18 (control for RNA loading in wells), and ribosomal 28S and 18S RNA bands are shown on the left and right sides, respectively. Animals with -416FIXm/1.4 and -679FIXm1 showed tissue specific expression patterns similar to that of -416FIXm1 (A) (data not shown). Interestingly, liver expression of hFIX observed for minigenes lacking the region containing AE5′ (except -770FIXm1, which remains to be tested as sufficient progeny animals become available) was high, but not as robust, as that seen with the natural gene. In addition, these minigenes expressed not only in the liver, but also in other tissues, such as kidney, lung and muscle, at various levels as high as ~20% of the liver level (Figure 7A). In clear contrast, animals with -802FIXm1 showed substantially liver-specific hFIX expression similar to that for the natural FIX gene (Figure 7B). These results suggest that the AE5' region controls liver specific expression of hFIX.

An apolipoprotein(a) transcription cotnrol region (ACR) whcih contains an ETS family target sequence 5'-CCCGGAAG-3' (SEQ ID NO:48) has been shown to exhibit enahncer activity *in vitro* in liver-derived HepG2 cells. However, the ACR does not appear to be liver-specific [Yang *et al.* (1998), *supra*].

EXAMPLE 6

Expression Of PEA-3 Protein In HepG2 Liver Cells

Expression of the transgene FIX was observed *in vivo*, but not *in vitro* in HepG2 cells, when expression vectors containing AE5' were used (See, Examples 2 and 3, *supra*). This observation, together with the absence of a footprint in HepG2 cell NEs (See, Example 4, *supra*) suggested to the inventors that HepG2 cells' lack of expression of the FIX transgene may be a result of the cells' expression of low levels of the PEA-3 protein (and/or the PEA-3 related protein) which binds to homologs of the invention's PEA-3 nucleotide sequence. The complete human PEA-3 cDNA has not yet been cloned (the human PEA-3 cDNA sequence of GenBank accession number U18018 lacks 8 amino acids at the N-terminal region when compared to the mouse PEA-3 cDNA sequence). In order to determine the role of the PEA-3 nucleotide sequence in gene expression *in vitro*, HepG2 cells which overexpress mouse PEA-3 protein were constructed as follows.

Expression constructs containing the mouse PEA-3 cDNA sequence (GenBank Accession Number X63190; Figure 9) were constructed as follows. Using the reported mouse PEA-3 cDNA sequence three sets of PCR primers were synthesized such that the entire coding region and parts of the flanking sequences would be amplified. Reverse transcription PCR (RT-PCR) was carried out, and the amplified mouse PEA-3 cDNA sequence was inserted into an expression vector under the control of the SV40 promoter, which does not interfere with the factor IX promoter (data not shown).

The PEA-3 expression vector is used to transfect HepG2 cells using the FuGene 6 (Boehringer Manheim) since this method was shown to improve transfection efficiency (See, Example 2, *supra*). Transfected HepG2 cells are screened for expression of PEA-3 by Northern blot analysis and/or Western blot analysis using commercially available antibody. Transfected HepG2 cell lines which stably express PEA-3 protein are selected for further use, *e.g.*, to analyze the underlying mechanism of PEA-3 action.

EXAMPLE 7

In vitro And *In Vivo* Expression Of Exemplary Human Protein C Minigene Expression Vectors Containing AE5' And AE3'

Protein C is a factor which plays a critical role in the anti-blood coagulation mechanism. Unlike factor IX, whose level in the circulation substantially increases with advancing age, protein C levels in the circulation do not increase with advancing age, but rather show a slight decrease over time. This decrease in circulating protein C levels is believed by the inventors to be the result of regulation at the gene transcription level. For this reason, the protein C gene provides an interesting exemplary gene for demonstrating the universality of the AE5' and AE3' function in gene expression both *in vitro* and *in vivo* as follows. In this Example, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594].

A. Construction of Human Protein C Minigene Expression Vectors

The human protein C genomic sequence has been previously reported (GenBank accession number M11228; Figure 12B]. Using this sequence, three protein C minigene expression vectors were prepared. The first human protein C minigene vector (-1426PCm1)

contained the human protein C promoter region of the protein C gene (GenBank accession number M11228; Figure 12 B) ligated to the human protein C cDNA (GenBank accession number X02750; Figure 12 A) which contains the first entire intron and poly-A sequence. The second human protein C minigene (AE5'/-1426PCm1) was the same as the first vector except that it additionally contained the nucleotide sequence AE5' at the 5' end of the human protein C cDNA. The third human protein C minigene (AE5'/-1426PCm1/AE3') was the same as the first vector vector except that it additionally contained the nucleotide sequences AE5' and AE3' at the 5' and 3' ends, respectively, of the human protein C cDNA.

B. Transient Expression Of human Protein C *in vitro* In HepG2 Cells

Each of the protein C minigene expression vectors was transfected into HepG2 cells using the FuGene 6 (Boehringer Mannheim). Four to five independent assays of human protein C activity were carried out as previously described [Turkey et al (1999) Throm. Haemost. 81:727-732]. HepG2 cells transfected with the -1426PCm1 vector showed *in vitro* transient protein C activities which were comperable to the activities shown by HepG2 cells transfected with the AE5'/-1426PCm1 vector (*i.e.*, (60-70 ng/10E6 cells/24 hrs).

C. Generation Of Transgenic Mice Harboring The protein C Minigene Expression Vectors

In order to determine whether AE5' in combination with AE3' is capable of increasing human protein C expression with advancing age, as observed for factor IX expression (Example 3, *supra*), transgenic mice which harbor the protein C minigene expression vectors are generated according to standard methods [Hogan et al. (1994), *supra*]. Briefly, protein C minigene vector plasmids are injected into fertilized eggs of C57B/6 X SJL mice and implanted into foster mother animals (CD-1). Offspring produced are screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples using two pairs of primers which are designed to amplify a unique fragment from the protein C transgenes and a 494 bp fragment from the mouse β-globin gene, respectively.

Founders are back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals are generated by crossing among heterozygous

F1 littermates and the following generations are similarly generated. Zygosity status of animals is determined by quantitative multiplex PCR analysis as described above. Founder lines for each minigene construct are subjected to longitudinal analysis for their entire life spans up to two years.

Circulatory human protein C levels are monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying the protein C minigene transgenes. Age-regulation patterns of circulatory human protein C levels are compared among all animals for each specific construct, different founder lines, different initial human protein C level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, and zygosity (homozygous / heterozygous) status of the transgenes.

Northern blot analyses of the liver RNA samples from animals is carried out using stringent washing conditions to determine whether any changes in circulatory human protein C levels are correlated with similar changes in the steady-state liver human protein C mRNA, rather than with loss of the human protein C transgene with age or with changes in human protein C turnover time. Observation of transgenic animals which contain the human protein C sequence as well as AE5' and AE3' and which increase stable circulatory human protein C levels with increasing animal age, as compared to the levels in transgenic animals which express the human protein C sequence in the absence of AE5' and AE3', demonstrates that the combination of AE5' and AE3' functions in age-stable expression of the exemplary human protein C gene. This observation will confirm that these results may ba achieved for genes other than the exempary hFIX and protein C genes.

EXAMPLE 8

In Vivo Expression Of Exemplary Expression Vectors Containing The Cytomegalovirus (CMV) Promoter And The AE5' And AE3'

This Example is carried out to demonstrate the universality of the age-related gene expression regulatory function of AE5' and AE3' with viral promoters. The CMV promoter is currently used in several gene therapy constructs but its activity decreases with time in the Liver. Furthermore, the activity of the CMVpromoter in the liver of transgenic mice is known to be lower than the activity in other tissues, such as muscle. Thus, this Example

investigates whether the combination of AE5' and AE3' halts or reverses the decline in the activity of the CMVpromoter in the liver.

The above-described -416FIXm1 expression vector (Example 1, and Figure 1) is used to construct a control vector to determine the effect of AE5' and AE3' on liver and circulatory levels of expression of human factor IX in transgenic animals. The control vector in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3' is constructed by replacing the human factor IX promoter sequence with the CMV promoter sequence (National Vector Core for Non-Viral Vectors at the University of Michigan) (the CMV promoter is also located between positions +1 to +596 in vector plasmid pCR3 from Invitrogen). The resultant expression vector in which the human factor IX gene is under the control of the CMV promoter is transfected into HepG2 cells. Transfected cells are expected to show human factor IX activity.

The -802FIXm1/1.4 vector of Figure 1 is used to construct a test vector in which the human factor IX promoter sequence of -802FIXm1/1.4 vector (which contains both AE5' and AE3') is replaced with the CMV promoter sequence.

In order to determine whether the combination of AE5' and AE3' is capable of increasing human factor IX expression with advancing age under the control of the CMV promoter, as observed for factor IX expression under the control of the factor IX promoter (Example 3, *supra*), transgenic mice which harbor either the control vector or the test vector are generated according to standard methods as described supra (Examples 3 and 7). The mRNA levels of human factor IX in the blood, liver and other tissues are monitored during longitudinal analyses of the transgenic mice. Age-regulation patterns of human factor IX mRNA levels in the different tissues are compared among all animals as described *supra* for each specific construct, different founder lines, different initial human factor IX levels at pre-pubertal age due to transgene positional effects, generation, sex, and zygosity status of the transgenes.

The observation of transgenic animals which contain the test vector and which increase stable circulatory human factor IX mRNA levels, as compared to the circulatory mRNA levels in transgenic animals which contain the control vector, demonstrates that the combination of AE5' and AE3' functions in increasing the activity of the exemplary CMV promoter.

EXAMPLE 9

Liver-Specific Expression Of The Exmplary Human Factor IX Gene Under The Control Of The CMV Promoter

This Example investigates whether the presence of AE5' imparts liver specific activity to the CMV promoter, which otherwise drives gene expression in several tissues in addition to the liver.

The -802FIXm1 vector which contains AE5' and lacks AE3' (Figure 1) is used to construct a test vector in which the human factor IX promoter sequence of the -802FIXm1 vector is replaced with the CMV promoter sequence. This test vector is used in parallel experiments with the control vector of Example 8 in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3'. Northern blot analysis is carried out as described *supra* (Example 3) in transgenic mice carrying the control vector or test vector. Animals expressing hFIX at high level are sacrificed at one month of age and total RNA is extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The levels of hFIX mRNA in the different tissues are compared. It is expected that transgenic animals harboring the control vector will express hFIX mRNA in liver as well as in at least one other tissue. In contrast, the observation that transgenic animals which harbor the test vector express hFIX mRNA in the liver and not in other tissues indicates that AE5' confers liver-specific activity to the exemplary CMV promoter.

From the above, it is clear that the invention provides methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

EXAMPLE 10

Construction Of A Series of Exemplary Human Protein C Minigene Expression Vectors

This Example and the following Examples 11-12 were carried out to further demonstrate the universality of the function of AE5' and AE3'' sequences with respect to regulating expression of the exemplary human protein C as described in Example 7, *supra*. The human protein C genomic sequence has been previously reported (GenBank accession

number M11228; Figure 12B]. In particular, Figure 14 shows the nucleotide sequence (SEQ ID NO:85) of the 5'-end of the human protein C gene [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. In Figure 14, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. Two minor start sites are marked with double asterisks. Exons are underlined. The translation start codon (ATG) is shown in boldface.

Using this sequence, eight protein C minigene expression vectors were prepared as shown in Figure 15. Fat lines at 5' and 3' ends represent flanking sequences. Introns are shown by blank rectangles. Exons are shown by shaded rectangles (corresponding to the 3' UTR part of the last exon). Arrows indicate transcriptional start sites. The vectors contained the human FIX AE5' sequence (SEQ ID NO:1; nt -802 through -771 of Figure 8; 32 bp in size) linked to Sph I linker sequences and/or the human FIX AE3'' sequence (SEQ ID NO:93; nt 32,110 through nt 32,263 of Figure 8, 154-bp long) linked to Sse8387 I linker sequences. All the amplified sequences of the human protein C minigene vectors were verified by dideoxy sequencing.

The first human protein C (hPC) minigene vector (-1462hPCm1) was composed of the 5' flanking sequence of human protein C up to nt -1462, exon I sequence, the complete first intron at the natural site (1431 bp in length) and the contiguous following sequence containing exons 2-9 (derived from the hPC cDNA) and the 3' immediate flanking genomic sequence through nt 11,108 (325 bp in length). As shown in Figure 15, nucleotide 67 was the first base of intron I, nucleotide 1,497 was the first base of exon 2 (the first base of the Met codon is nucleotide 1,514), and nucleotide 10,488 was the last base of the stop codon.

To construct -1462hPCm1, a region spanning nt -1462 through 1,560 (3,022 bp in length) of the hPC gene was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) using 5' and 3' primers containing Sph I linker and the unique internal Msc I site in the exon 2, respectively, and human genomic DNA as the template. The generated fragment, containing the 5' flanking sequence, exon I, intron I and a short 5' portion of exon 2 to the internal Msc I site, was then inserted into a hPC cDNA plasmid, PUC119-hPC, in between Sph I and Msc I sites by replacing its 5' portion of the hPC cDNA sequence. The 3' end of the resulting minigene, -1462hPCm', contained the entire 3' UTR, but only up to poly(A) attachment site (nt 10,783). Its 3' end region (the 3' sequence beyond the internal Sse8387 I site in the 3' UTR) was then freed by Sse8387 I/EcoR I

double digestion, and replaced with a Sse8387 I/EcoR I fragment (612 bp in the length, spanning nt 10,497 through 11,108 of hPC gene), which was generated by PCR. The hPC minigene, thus constructed was named -1462hPCm1, (approximately 4981 bp in length) and served as a parent construct for generating other hPC minigene constructs for generating other hPC minigene constructs as described below. This hPC minigene vector was used to construct the first transgenic mouse colony, which was used as a positive control for hPC expression.

The second human protein C minigene vector (-82hPCm1) was another control human hPC minigene. It was the same as -1462hPCm1, except that its promoter region was extended only up to nt -82 instead of to nt -1462, and thus contained the hPC upstream sequence (SEQ ID NO:86) from nt -82 to nt +1. Minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR spanning nt -82 in the 5' upstream through the internal Msc I site, thus generating -82hPCm1.

The third human protein C minigene vector (AE5'/-1462hPCm1) was the same as -1462hPCm1, except that it had an AE5' sequence (32 bp in length) inserted at the 5' end at an Sph I site.

The fourth human protein C minigene vector (-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had an AE3'' sequence inserted at the Sse8387 I site within the 3' untranslated region (UTR) of the hPC minigene.

The fifth human protein C minigene vector (AE5'/-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had both AE5' and AE3''. The AE5' fragment (32 bp in length, spanning nt -802 through -771 of hFIX gene) was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) with human genomic DNA as template and PCR primers containing Sph I linker sequences. The AE3'' fragment spanning nt 32,110 through nt 32,263, and containing the potential stem-loop structure forming region (nt 32,142 through 32,243) was produced by PCR using primers containing the internal Sse8387 I site sequences. AE5' and AE3'' fragments with Sph I or Sse8387 I sticky ends, respectively, were then inserted into -1462hPCm1 at Sph I site at the 5' end and Sse8387 I site in the 3'

UTR, respectively, thus generating minigene AE5'/-1462hPCm1/AE3''. Therefore, AE5'/-1462hPCm1/AE3'' is the same as -1462hPCm1 except it has both AE5' and AE3' sequences of the hFIX gene at the 5' end Sph I site and Sse8387 I site in the 3' UTR, respectively.

The sixth and seventh human protein C minigene vectors (-849hPCm1 and -802hPCm1) were the same as -1462hPCm1, except that their 5' end sequences extended to nt -849 and -802, respectively, instead of to -1462. Construction of these minigenes was essentially the same as construction of the second human protein C minigene (-82hPCm1); minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream region through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR, spanning nt -849, or -802 in the 5' upstream through the internal Msc I site, thus generating -849hPCm1 and -802hPCm1, respectively.

The eighth human protein C minigene vector (AE5'/-82hPCm1) was generated by inserting the AE5' sequence with Sph I sticky ends into the second vector, -82hPCm1, at the 5' end Sph I site.

EXAMPLE 11

Transient Expression of Exemplary Human Protein C Minigene Expression Vectors *In Vitro* In Human Hepatoma HepG2 Cell Line

Transient expression activities of the hPC minigenes of Example 10 were tested with HepG2 cells as described in Example 2, *supra*. Figure 16 shows the relative *in vitro* transient expression activities and standard deviations of the human protein C minigene expression constructs relative to the activity (approximately 70 ng/10^6 cells/48 hr, defined as 100% activity) of the control -1462hPCm1 construct.

The relative transient expression activities in Figure 16 show that the results obtained with AE3'' when in tandem with the hPC gene are fully consistent with what the inventors previously observed with AE3'' when in tandem with the hFIX gene (Figure 1); the presence of AE3'' showed approximately a 30% suppression in transient expression in comparison to the minigenes which lacked AE3'' (*i.e.*, -1462hPCm1/AE3'' compared with

-1462hPCm1, and AE5'/-1462hPCm1/AE3'' compared with AE5'/-1462hPCm1). The control constructs -1462hPCm1 and -82hPCm1 showed similar transient expression activities to each other.

As discussed above, these results were surprising because they were contrary to those previously reported by Miao et al. (1996), *supra*, when using a heterologous reporter gene, chloramphenicol acetyltranferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences.

EXAMPLE 12

Generation And Analysis Of Transgenic Mice Harboring Human Protein C Minigene Expression Vectors

Transgenic animals were constructed using each of the eight expression plasmids described above in Example 10 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition) as described in Example 3, *supra*. Fertilized eggs of mice were microinjected with the minigene transgene DNA and implanted into foster mother animals.

Circulatory hPC levels were monitored during longitudinal analyses of transgenic mice harboring the -1462hPCm1, -82hPCm1, and AE5'/-1462hPCm1/AE3'' minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection via tail-tip snipping, and the obtained serum was routinely used to quantify hPC levels in the circulation using ELISA for each age point. The ELISA assay employed a mouse monoclonal anti-hPC antibody (Celsus Laboratories) as a first antibody and a rabbit polyclonal anti-hPC antibody (Celsus Laboratories) as a second antibody. Pooled human plasma (George King Bio-Medical) was used to prepare a hPC standard curve for each assay. The results are shown in Figure 17. The labeling in Figure 17 reflects the tag numbers of animals containing each minigene construct. Figure 17 shows representative animals with -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3'' (C) expression vectors.

Importantly, Figure 17 shows that age-regulation patterns were remarkably similar among all animals for each specific construct. In particular, the results show that transgenic animals containing the -1462hPC m1 construct contained age-stable levels of human protein C, *i.e.*, the animals expressed relatively constant levels of human protein C at different time

points during the life span of the transgenic animals (Figure 17A). In direct contrast, the presence of AE5' and AE3' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of AE5' and AE3'' sequences in regulating expression of operably linked genes in an age-related manner.

The data also shows that, whereas transgenic animals containing the -1462hPC m1 construct exhibited relatively constant and relatively high levels (from about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCm1 construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) of human protein C which declined at a precipitous rate over time. Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCm1 construct. These results demonstrate that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene directs age-stable expression as well as relatively higher levels of expression (as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.